THE RE-SKILLING REQUIREMENTS OF DATABASE ADMINISTRATORS IN A BIG DATA ENVIRONMENT

by

ROMNEY ELDRIDGE CHARLES DAVIDS

195064720

Thesis submitted in partial fulfilment of the requirements for the degree

Master of Technology: Information Technology

in the

Faculty of Informatics and Design

at the

Cape Peninsula University of Technology

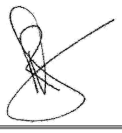
Supervisor: Mr Jay Barnes

Cape Town

Date submitted: 11 December 2018

DECLARATION

I, Romney Eldridge Charles Davids, declare that the contents of this thesis represent my own unaided work, and that the thesis has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.



11 December 2018

Signed **Date**

ABSTRACT

Big data generated within organisations is on an exponential increase. It is not only the increase in the volume and velocity of structured data but more so the complexity of unstructured data that has presented new challenges. Organisations have since realised the potential to derive new business value from this unstructured data and therefore, the requirement to manage such data has arisen. Even though the traditional database administrator (DBA) administers the organisation's traditional structured data within the relational database management system (RDBMS) environment, the unstructured nature of big data presents potential new challenges. As a result, there is a requirement for new technologies to store and process both structured and unstructured data, as well as a range of skill sets that are required to administer it. It is however not clear whether these skill sets exist within a single individual such as the traditional DBA or within different roles across various functional departments within the organisation. This study explores the difference in skill set between the DBA administering the RDBMS environment and the requirement to transition into a similar position within a big data environment. It further compares the current DBA's skill set with other roles found within the big data environment to explore where the DBA role fits into the big data environment.

Keywords: Big data, database administrator (DBA), relational database management system (RDBMS), re-skilling

ACKNOWLEDGEMENTS

I wish to thank:

- My academic supervisor, Mr Jay Barnes for his continuous support, academic guidance, professionalism and really going the extra mile for his students.
- Dr Andre de la Harpe for his support and academic guidance.
- CPUT for allowing me the opportunity to contribute towards research within the Faculty of Informatics and Design.
- The organisations who made their resources available to take part in this study including the respondents who sacrificed their time to take part.

DEDICATION

I dedicate this thesis to my loving wife, Bonita Davids and daughters Jamie-Lee and Carly Davids. Without them this would not have been possible. Their continuous support, words of encouragement and patience carried me through. The many sacrifices they made to support me lifted my spirit and gave me strength to pursue the end goal. Without them, this would not have been possible. Thank you!

My parents, Mr & Mrs Davids and Mr & Mrs van Aarde, for the constant support and prayers which gave me strength. Your support and guidance are appreciated. Thank you!

“I can do all things through Christ which strengtheneth me.”
Philippians 4:13 (KJV)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Abbreviation	Description
ACID	Atomicity, Consistency, Isolation, Durability
ANSI	American National Standards Institute
AP	Available and partition-resilient
BASE	Basically Available, Soft State, Eventually Consistent
BLOB	Binary Large Object
BSON	Binary JSON
CLI	Command-line Interface
CP	Consistent and partition-resilient
CS	Computer Science
CSV	Comma-Separated Values
DBA	Database Administrator
DBaaS	Database-as-a-Service
DBMS	Database Management System
DCL	Data control language
DDL	Data definition language
DevOps	Development and Operations
DML	Data manipulation language
ETL	Extract-Transform-Load
GFS	Google File System
HDFS	Hadoop Distributed File System
I/O	Input-output
IaaS	Infrastructure-as-a-Service
IP	Intellectual Property
IT	Information Technology
JSON	JavaScript Object Notation
MB	Megabyte
MS	Microsoft
NBD-PWG	NIST Big Data Public Working Group
NBDRA	NIST Big Data Reference Architecture
NewSQL	NewSQL
NoSQL	Not only SQL / Not Relational
PaaS	Platform-as-a-Service
RDBMS	Relational Database Management System
SaaS	Software-as-a-Service
SQL	Structured Query Language
XML	eXtensible Markup Language
YARN	Yet Another Resource Negotiator

CHAPTER ONE: INTRODUCTION

1.1 Introduction

The ever-changing Information Technology (IT) landscape has resulted in a move from a paper-based society to that of a digital environment. The transportation and generation of information now live in computer systems and the internet. The need to store all this information is on an exponential rise, resulting in everything being stored and very little if anything, being discarded (Kaisler, Armour, Espinosa & Money, 2013). This has subsequently put pressure on computer systems in terms of capacity and processing power so as to make this information available, as “data volume is scaling faster than computer resources” (Almeida & Calistru, 2013:16). As a result, the emergence of cloud-based systems with extensive storage capacity and processing power and the ability to scale on-demand has come to the fore to meet some of these information needs.

The need to accommodate the increased volumes of data has also become evident within many organisations. They have long since become aware that there is an opportunity to get value out of the information sent through interactions on the internet and company websites. Weblogs showing navigation paths through e-commerce websites have become a valuable source of information, allowing companies to focus their attention on the potential needs of their customers through their browsing behaviour and thus provide the potential for new revenue streams.

All of this has resulted in the phenomenon known as “big data”, a coming together of different data types and the requirement to store all this information. New technology requirements facilitating the storage of both structured and unstructured data have resulted in a move towards newer technologies to store unstructured data in conjunction with the traditional RDBMS (relational database management system) (Sharma, 2013). This move has resulted in an entirely new technology stack to accommodate this need. Database systems are now being implemented on networked clustered systems, making use of parallel processing with data being spread across each node in the cluster (Borkar, Carey & Li, 2012a). Non-relational database technologies such as NoSQL (i.e. Not Only SQL or Not Relational) and distributed processing frameworks such as Hadoop are now becoming popular.

As a result, it is now a requirement of the database administrator (DBA) to focus attention not only on the management and administration of the RDBMS but to shift to a more diverse environment, that of the big data environment, to meet the needs

of organisations. A new skill set is required in addition to the current skills of the DBA, in order to administer this new big data environment. These new skills overlap with those of other roles such as system administrators, data analysts and developers (Simmonds, 2013). A further complication is the rise of the data scientist, a new role within the big data environment (Davenport & Patil, 2012). It is far from clear whether traditional DBAs will continue in the role in which they are accustomed or if they will be required to gain a broader skill set to manage the new technologies introduced in a big data environment.

1.2 Background to the research problem

Data within organisations is continuously expanding in volume and becoming increasingly unstructured, making it increasingly difficult to continue managing and analysing it through traditional means (Davenport, Barth & Bean, 2012). The complexity and variety in which the data presents itself have also changed (Almeida & Calistru, 2013). Kaisler *et al.* (2013) suggest that issues such as storage, management and processing of data need to be re-considered, and that existing employees might not have the required skills and tools necessary to manage effectively this increasing stream of data. As a result, it is becoming challenging for traditional data management and processing tools to handle the size and complexity of the datasets within a big data environment (Hunter, 2013; Hashem, Yaqoob, Anuar, Mokhtar, Gani & Khan, 2015).

With the lowering costs and increase in storage capacity of computer systems, it is no longer standard practice to ensure that only the most important data is kept and the rest discarded (Kaisler *et al.*, 2013). As a result, data is being generated and stored from an extraordinary number of sources and at a tremendous rate. This ability to store and process increasing amounts of data can also be attributed to the increase in machine power to compute and analyse this data (Dijcks, 2011; Bloem, van Doorn, Duivestein, van Manen & van Ommeren, 2012).

Big data is a moving target as described by Kaisler *et al.* (2013). For some time, the increased volume of data has been counteracted by providing more processing power and storage capacity, but with big data, “there is a fundamental shift underway now considering that data volume is scaling faster than computer resources” (Almeida & Calistru, 2013:16). Traditional database systems can no longer manage the volumes and types of electronic information being generated daily (Courtney, 2013) and thus is unable to handle the requirements of big data (Storey & Song, 2017). There is thus a requirement for organisations to improve their IT

infrastructures in order to integrate the surge of big data along with their traditional enterprise data (Dijcks, 2011). Borkar, Carey and Li (2012b) note that in the year 2000, renewed attention was given by companies like Google, Facebook, and Yahoo! to develop new technologies to accommodate the big data phenomenon driven by web-scale data. Organisations can only capture the value of big data by investing in new storage mediums, technologies and techniques by which to store and analyse the data (Manyika *et al.*, 2011). Davenport *et al.* (2012) agree that these major improvements are necessary to keep up with the increasing volumes of data in big data environments. These factors have also resulted in database systems having to evolve in order to accommodate big data (Holt, Ramage, Kear & Heap, 2015).

Davenport *et al.* (2012:24) note there is an "information ecosystem" developing within the organisation through the continuous exchange of information, both internal and external to the organisation, which has resulted from the continued evolution of big data. As a result, the value of digitised consumer data within the organisation has also increased, thus requiring particular attention to be given to policies pertaining to the safe-keeping of such information (Almeida & Calistru, 2013).

In order for organisations to remain competitive, they will need to explore the potential value that big data holds, as many leading companies have already started to explore this new avenue (Almeida & Calistru, 2013). In 2011 Gartner (cited in Bloem *et al.*, 2012) suggested that organisations that could consolidate their diverse information sources into a central information management system could potentially gain a competitive advantage of over 20%. It is vital for organisations to have the correct set of skills and expertise, alongside the technology infrastructure and processes, to exploit such a potential advantage through the field of data science.

Manyika *et al.* (2011) noted that there were "5 billion mobile phones in use in 2010, 30 billion pieces of content shared on Facebook every month and a 40% projected growth in global data generated per year vs 5% growth in global IT spending". They also noted that there would be a requirement for approximately 140,000 to 190,000 jobs for individuals to perform deep analysis on big data and 1.5 million managers with the ability to analyse and make use of this data in the United States alone. Asthana (2011) believes that a new role with a different skill set is needed, namely a data scientist. People occupying this role must have a wide variety of skills including technical IT skills, business, and process knowledge within the organisation in addition to the analytical skills required to interpret big data (Davenport *et al.*, 2012; Davenport & Patil, 2012). They must also be able to write code, have strong

computational skills and the ability to extract meaningful value from the mountain of unstructured data. It is not clear whether this skill set is within the capabilities of a single individual or a team of specialists (Australian Government Department of Finance and Deregulation, 2013).

Because of the structural complexity and volumes of data within big data, newer technologies that can process such data are required. No single technology can accommodate the requirements of big data. There are a variety of solutions, each attempting to solve a different problem and this in turn also make it difficult to determine the skills required by organisations (Volk, Bosse & Turowski, 2017).

All these new requirements do not do away with the need for traditional technologies such as the RDBMS administered by the traditional DBA, as it is still best suited to process structural data. There is a need for both older and newer technologies to complement one another in order to successfully implement a big data solution (Sharma, 2013). The traditional DBA's skill set is therefore still required as an integral requirement, in addition to the skills required to administer newer technologies housed within a big data environment.

1.3 Research problem statement

There is currently a shortage of specialists to administer big data environments in organisations. Even though DBAs manage the organisations traditional RDBMS, they do not have the required skills needed to work in a big data environment. Organisations might therefore not be able to extract the potential value that big data holds if they do not have the necessary skilled staff to manage and analyse big data within the organisation.

1.4 Research questions, approach, aims and objectives

Table 1-1: Main research question and sub-questions

Research main question		What is the skills gap between the DBA working in an RDBMS environment and a similar role in a big data environment?	
Research sub-questions (RSQ)		**Research Approach**	**Objectives**
RSQ1	What are the current skill set and responsibilities of a DBA working in an RDBMS environment?	Interviews	To explore the current skill set and responsibilities of DBAs working in an RDBMS environment.
RSQ2	What are the skills and technology requirements of a big data environment?	Interviews	To explore what are the skill set and technology requirements imposed by a big data environment.
RSQ3	What career plans do organisations have for DBAs with regards to big data management?	Interviews	To explore what career plans organisations have in place to enable them to transition towards a role in a big data environment.

1.4.1 Research aim

This study aims to explore the difference in skill set between a DBA administering an RDBMS and similar position in a big data environment.

1.4.2 Research objectives

The objectives of this study are to:

- explore what technical re-skilling will be required of DBAs when transitioning from a position within an RDBMS environment to that of a big data environment.
- develop a model to assist organisations in accommodating the DBA role within the big data environment.

1.5 Research methodology

1.5.1 Research philosophy

When conducting research, different philosophical paradigms must be considered, such as subjectivism and objectivism, and these paradigms are defined by different assumptions. These assumptions relate to how the researcher experiences the world and are influenced by personal believes and values (Remenyi, Pather & Klopper, 2011). There are two core assumptions in this regard; that is what constitutes reality (ontology) and how knowledge of reality is acquired (epistemology).

Ontology can be defined along two different ways of thinking. The first dictates that the nature of reality exists independently of human influence (objectivism) while the alternate acknowledges that there is human influence (constructionism or subjectivism or interpretivism). Like the ontological paradigm, epistemology can be viewed from a positivist or interpretivist paradigm. The positivist dictates that the study of reality must be conducted scientifically (Gray, 2013:20) as knowledge is generated through empirical evidence and must be repeatable (Wahyuni, 2012). Alternatively, the interpretivist considers that knowledge of reality is defined through the subject's perspective and this is often influenced by their experiences and backgrounds (Wahyuni, 2012).

It is reasonable to assume that none of the DBAs' working environments will be the same and thus their experiences of the real-world including technologies used will be different. Therefore, in relation to both ontological and epistemological assumptions, there is influence and participation from the social actors in this study (DBAs) about what constitutes their reality including how knowledge of that reality is acquired. With

this in mind, this study will adopt constructionist ontological and interpretivist epistemological stances.

1.5.2 Research approach

Many studies relating to information systems (IS) are conducted using a qualitative methodology, and it takes into consideration the interactions of people with such systems (Kaplan & Maxwell, 2005). Qualitative research considers the experiences of the participants to understand the complexities present (Tuli, 2010). Quantitative research, on the other hand, dictates that the outcomes of research be repeatable. It is conducted through scientific enquiry with a reliance on measurements while qualitative studies are "measured" through the researcher's interpretation (Stake, 2013). The DBAs' environment, technology, is continually changing, and the experience of participants will differ due to the different environments they find themselves in. This changing environment does not meet the requirements of a quantitative study. This study will thus follow a qualitative approach, and this is in line with the constructionist ontological, and interpretivist epistemological stances adopted.

There are no pre-existing theories upon which this research is based, and the area of concern will be explored through the data that emerges from the study. This study is therefore not attempting to test or confirm existing theories which are the basis of a deductive approach (Gray, 2013:16). A deductive approach is also largely associated with quantitative research (Atieno, 2009). Qualitative research allows theories to emerge from the data and is inductive by nature (Kaplan & Maxwell, 2005), and hence this study will follow an inductive approach.

1.6 Research design

This study aims to explore the difference in skill set between a DBA working in an RDBMS environment and a similar position within a big data environment. This study follows an inductive approach and relies on qualitative data. Exploratory research often uses qualitative data (Ellis & Levy, 2009), allowing enquiry into an unknown area while providing the researcher with a certain degree of freedom through their interpretation of the findings and building towards potential theories from the data that emerges (Jupp, 2006).

1.7 Data collection

Primary data will be collected from the participants in this study using one of the most common and accepted data collection methods employed within qualitative studies, namely interviews (Tuli, 2010).

1.7.1 Units of analysis

The subject from which data will be collected in a study is often referred to as the unit of analysis (UoA) (Yin, 2006). This study will focus on IT professionals who administer the organisation's databases, and such individuals commonly have the job title of "database administrator or DBA". They will be the UoA for this study.

1.7.2 Sampling

This study made use of non-probability convenience sampling. Non-probability sampling is an often-used sampling method within qualitative studies and is used when looking for individuals with specific skills or characteristics (Mays & Pope, 1995). DBAs match these characteristics as they are specialists within the IT sector with a specific skill set to administer databases.

It is not the intention of the study to generalise the findings to the general DBA population. Qualitative studies allow for smaller sample sizes, and 15 interviews will be targeted. This is considered an acceptable sample size (Kvale, 2011c:43-44). If data saturation is not achieved from these 15 interviews, more will be added as necessary.

1.7.3 Interviews

Interviews are one of the data collection methods used within qualitative studies (Kaplan & Maxwell, 2005). It is a form of conversation between the interviewer and the interviewee with the intention of generating knowledge (Kvale, 2011b:7). Semi-structured interviews will be conducted in this research because they provide insight into complex and unexplored areas through peoples' opinions and experiences (Louise Barriball & While, 1994).

1.8 Data analysis

1.8.1 Transcribing the interviews

Recording interviews allows the researcher to pay more attention to the interview and not be distracted by having to write (Kvale, 2011d:93). Where consent was provided, interviews were audio recorded and later transcribed verbatim by the

researcher. Transcription was not a once off process but required several iterations for each interview recording. The transcription was anonymised where participants referred to the names of companies or institutions. None of the audio contained any company intellectual property (IP), personal or sensitive information and the audio files were safeguarded in a password protected zip file. All audio recordings and interview transcripts are available on request (see APPENDIX D for two transcribed interviews).

1.8.2 Thematic analysis

The data analysis process adopted for this study was thematic analysis. This allows for "identifying, analysing, and reporting patterns (themes) within data" (Braun & Clarke, 2006:6). The process allows for themes to emerge from the raw data (Thomas, 2003) as no pre-existing theories are being tested.

The thematic analysis process is an iterative process. Initial keywords and key phrases were identified in the interview text and assigned a code. A semantic approach was used as it was not the intention to identify underlying factors to what was reported by respondents (Braun & Clarke, 2006). No pre-defined list of codes was used, and the researcher relied on knowledge of the literature and data collected to generate codes (Kaplan & Maxwell, 2005). Similar codes were grouped, broadening the level of analysis to categories. Each code within a category was given a strength indicator (N) which reported the number of respondents from which such codes appeared. Likewise, similar categories were grouped, and the analysis broadened towards themes. The themes were then used to identify patterns and ultimately aid the researcher in answering the research questions (Braun & Clarke, 2006) (see APPENDIX E for two coded transcripts).

1.9 Delineation of the research

This study focuses purely on the technical and not the business or social re-skilling requirements of DBAs. Semi-structured interviews were conducted face-to-face with respondents working in various sectors within South Africa. Re-skilling in other areas affected by the requirements of a big data environment will not be considered and can be investigated in future research.

1.10 Ethical considerations

Even though interviews allow for free conversation between the researcher and participant, it is important that the researcher be cognisant of the purpose of the research (Orb, Eisenhauer & Wynaden, 2001). Interviews can be an intrusive

process, and it is essential for the researcher to consider this (Kaplan & Maxwell, 2005), irrespective of the type of questions to be asked.

While consent will be sought from the organisation's management, Joe, Raben and Phillips (2016:6) consider it of utmost importance for the participant to take part voluntarily. Such consent includes providing the participant with enough information to be informed about the study, purpose, usage, and confidentiality of information collected, and importantly, anonymity. The following ethical considerations were adhered to in this study:

Informed consent: Respondents were briefed about the purpose of the study and information to be collected before the start of the interview. Respondents acknowledged their consent by signing an interview consent form (see APPENDIX B).

Anonymity: The researcher advised respondents to indicate the desire not to be named on any publications resulting from the study. Respondents were given a pseudonym and addressed as such for the duration of the interview. References to organisations and institutions were anonymised upon transcription of the interview.

Recording of interviews: Permission was sought to record interviews, and the verbal consent of the respondent was also recorded. This was done in full view of the respondent so that they were able to see the interviewer handle the recording device.

Voluntary participation: No incentives were given to take part, and respondents could withdraw at any stage. After each interview question respondents were asked if they wanted to continue with the interview and this was recorded.

Personal and sensitive information: None of the questions in the study related to any personal or sensitive information, company trade secrets, intellectual property rights or financial statements.

1.11 The research process

Below is a schematic representation of the research process applied within this study.

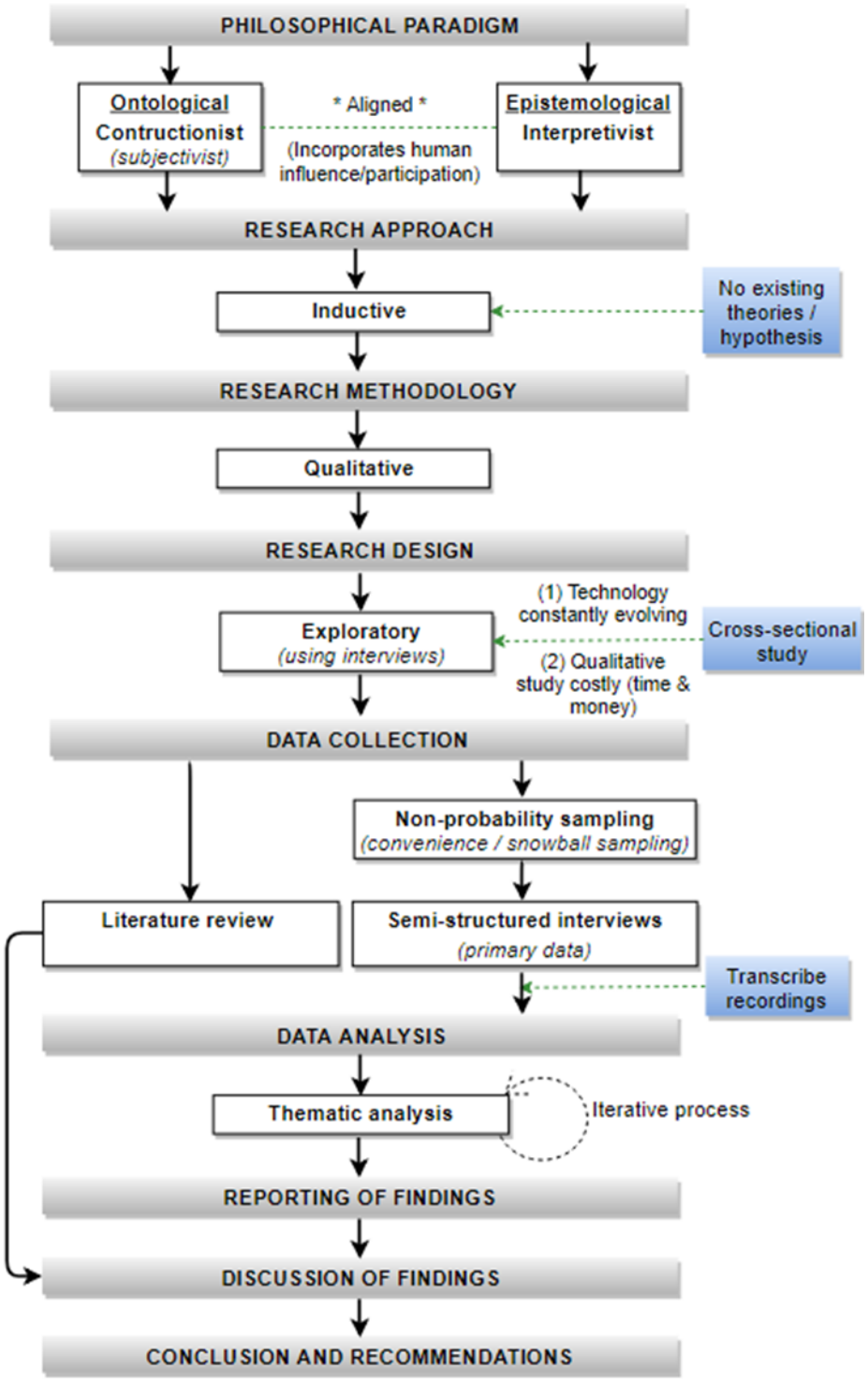


Figure 1-1: Research process followed in this study

1.12 Chapter outline

The thesis consists of eight chapters in addition to the preamble, references, and appendices sections. A brief description of each chapter will now be given:

Chapter One - Introduction: An introduction and background to the research problem are provided. These are accompanied by the research problem, research

question and sub-questions, aim and objectives, research methodology, research strategy, data collection, data analysis, delineation and ethical considerations.

Chapter Two - Literature review: An in-depth literature review of the RDBMS environment relative to its characteristics and current challenges faced. The role and characteristics of the DBA administering this environment are defined. Next, the big data landscape and its architecture are introduced. A high-level overview of some of the core technologies that supports a big data environment is presented to set the stage for the possible transition of the DBA to a big data environment.

Chapter Three - Research design and methodology: It presents the research design and methodology applied in this study based on the ontological and epistemological stances adopted. It describes the research approach (qualitative), research design (exploratory), data collection (interviews) and data analysis (thematic analysis) techniques selected, and the relationship between them. Ethical considerations and delineation within this study conclude the chapter.

Chapter Four - Data collection and data analysis: An actual account of the research process undertaken in this study. It presents how the fieldwork was conducted and how the data analysis process was applied to the data collected.

Chapter Five - Research findings: The results of the 15 interviews conducted in this study are reported. The results are presented through seven themes and the underlying categories that emerged from the analysis process.

Chapter Six - Discussion of findings: A detailed discussion of the seven themes as reported in the findings is presented. The discussion of the findings is based on the literature presented in this study (Chapter Two) in addition to newly acquired literature subsequently referenced to support the discussion.

Chapter Seven - Conclusion and recommendations: Answers the research main question and research sub-questions, based on the findings and literature. The researcher's contribution to the body of knowledge, recommendations and suggestions for future research concludes this chapter.

Chapter Eight - My research journey: An account of the researcher’s journey is presented in relation to the process undertaken and the obstacles encountered.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Chapter Two presents a high-level overview of the current RDBMS environment in relation to one of the primary technologies used, the relational database. It also presents a profile of the primary person responsible for managing the database, namely the DBA. Considering this research is intending to explore the skills required of the DBA occupying a similar role within the big data environment, the literature will also include an overview of the big data landscape. The intention is not to list all technologies and trends currently in place but to focus on specific aspects to show the difference between existing RDBMS technologies, and the changing role and focus of the DBA. The new big data technologies will also be used to show the change in environment in which the DBA finds themselves, and their need to adapt to these changing technologies regarding their current skill set.

2.2 The RDBMS environment

2.2.1 An introduction to the RDBMS

While database systems emerged as early as the 1960s, the RDBMS only started to take effect at the start of the 1980s after the initial relational model by E.F. Codd (Feuerlicht, 2010). The purpose of these databases was primarily to accommodate structured data (Atzeni, Jensen, Orsi, Ram, Tanca & Torlone, 2013).

Structured data is managed within a relational database as rows and columns in a table and is used by organisations to drive the organisational needs and business decisions (Blumberg & Atre, 2003). Data can be categorised in columns within tables that have fixed names, data types and layouts often with complex relationships amongst these tables (Arora & Gupta, 2012). These relationships between tables are modelled using constraints such as primary, foreign keys and indexes (Binani, Gutti & Upadhyay, 2016).

All aspects relating to the management of the database and structured data within an RDBMS is performed using a standard query language known as structured query language (SQL) (Binani *et al.*, 2016). These operations and functions are categorised within the SQL language as data definition language (DDL), data manipulation language (DML) and data control language (DCL) (Silva, Almeida & Queiroz, 2016).

2.2.2 Characteristics of the RDBMS

RDBMS conform to ACID guarantees in order to ensure the integrity of the transactional data, where the acronym of ACID stands for atomicity, consistency, isolation and durability (Leavitt, 2010; Binani *et al.*, 2016). The integrity of the data is thus maintained by making use of transactions when interacting with data in an RDBMS (Moniruzzaman & Hossain, 2013). The "atomicity" guarantee means that either an update, insert or delete statement is performed completely or it is rolled back (Leavitt, 2010; Binani *et al.*, 2016:43). The "consistency" guarantee ensures that all the database rules must remain intact after that transaction has completed. The "isolation" guarantee refers to the transactions that run independently of one another, and finally, the "durability" guarantee ensures that the changes effected to the data will persist (Leavitt, 2010; Binani *et al.*, 2016:43).

2.2.3 Challenges faced by the RDBMS

Even though the volume of data being stored has increased significantly, the real problem facing the RDBMS is not only the volume of data but how such data evolved to become structurally complex. These diverse datasets must be combined with various technologies to become meaningful and valuable information (Sukumar, Olama, McNair & Nutaro, 2013). Thus, the ability of RDBMSs to effectively deal with unstructured data has increasingly come under the spotlight. Storey and Song (2017:54) state that RDBMSs "are simply not capable of handling big data".

The complexity arises with the RDBMS in that all data needs to be converted to fit the structure of the database design. The challenge posed by the unstructured nature of certain data types within big data is to be able to conform to such rigid requirements (Leavitt, 2010). Limitations also arise with the requirement to scale upwards, needing more powerful and expensive hardware to achieve this goal. In addition to this, the programming language used to query the data within an RDBMS, namely SQL, can become quite large and complex depending on the information to be retrieved (Leavitt, 2010).

Lee, Tang and Choi (2013) looked at the potential alternatives such as NoSQL and XML based databases as compared to relational databases to accommodate medical data. It was their opinion that the rigid nature enforced by RDBMS did not lend itself to the flexibility required by the medical data whereby placeholders in the form of columns were required to be in place even though the majority of the time this data would not be present as a result of incomplete records and information entered.

2.3 The role of the DBA

The role the DBA is not new and is a specialised role found within the IT sector. Peshkar and Ghosekar (2015:41) report that DBAs are also in some instances referred to as a “Database Coordinator or Database Programmer”. The primary focus of someone fulfilling the role of a DBA is “to provide an environment that is both convenient and efficient to use in retrieving and storing database information” (Peshkar & Ghosekar, 2015:41). Holt *et al.* (2015:163) describe database management as “the administrative tasks associated with the storage, modification and retrieval of data held within a database management system (DBMS)”. Even though the role has only one specific focus, the range of duties performed, and the skills required are quite broad. These responsibilities do not only include supporting the operational database but can also include the management of the entire lifecycle of the database, ranging from its initial inception such as the design to the upkeep and maintenance thereof (Simmonds, 2013).

Given the importance of information and data to an organisation’s existence, it is easy to see the importance of the role fulfilled by the DBA. The DBMS administered by DBAs is in many instances a critical and core component of many organisations’ business and technology stack (Feuerlicht, 2010; Moore, 2010). The role of the DBA is so critical that in some industries it can have a financial impact if the database is not available (Polakowski, 2009). The importance of the role of the DBA is highlighted by Peshkar and Ghosekar (2015) who stated that the database is valued according to the importance of the data to the organisation which it contains.

The role of the DBA can be split into three areas of interest or specialisation. The first is a production type role which relates to database systems supporting live operations. The second is the development or application type role which incorporates the design, development and maintenance of applications. Lastly is the role of data administration, dealing with all aspects relating to the management and administration of data (Peshkar & Ghosekar, 2015).

2.3.1 Skills required by a DBA

The skills required to fulfil the responsibilities of the DBA are not necessarily generic but, in many instances are specialised to a specific vendor or product they support (Simmonds, 2013). Examples of such product vendors are Oracle, IBM, and Microsoft (Moore, 2010). This does not mean that DBAs focus only on a specific

vendor; they can either specialise or support multiple vendors. Holt *et al.* (2015) found that many organisations do not employ a single database technology from a single vendor, but rather use multiple vendors and database technologies.

It is critical that DBAs firstly have the required technical skills related to the database software they support, including being proficient in SQL and how their data is modelled (Polakowski, 2009). The role of the DBA not only requires specific knowledge and expertise of the database but also demands a broad range of technical skills and knowledge (Holt *et al.*, 2015). These can include skills not directly related to the database software such as knowledge of the operating system and programming (Polakowski, 2009). Skills required by a DBA also include non-technical or soft-skills such as interpersonal and communication skills (Polakowski, 2009). Holt *et al.* (2015) found that their respondents reported communication as a key component of the DBA's skill set given that communication occurs across many different stakeholders.

2.3.2 Responsibilities of a DBA

"A database administrator is typically responsible for efficient hardware maintenance, logical database implementation, and security" (Litecky, Aken, Ahmad & Nelson, 2010:79). At a high level, Simmonds (2013) categorised these responsibilities as database design, database security and authorisation, ensuring data integrity, ensuring availability, performance monitoring and tuning, governance and regulatory compliance, and backup and recovery. Peshkar and Ghosekar (2015:41) similarly define the role as being "responsible for the installation, configuration, upgrade, administration, monitoring and maintenance of databases in an organisation".

DBAs do not work in isolation but work closely with the database developer and the database architect as a database team. The role also requires DBAs to work with other teams such as software developers, given their expertise relating to the database. DBAs provide software developers with guidance and assist with things such as SQL performance tuning and troubleshooting, and database objects that need to be designed or created (Moore, 2010). The DBA is pivotal to the successful functioning of this database team, ensuring that the database is always functioning and available (Sukumar *et al.*, 2013).

2.3.3 Formal qualifications of a DBA

Industry expects that DBAs will have some form of tertiary qualification, which would either be in a field such as Computer Science (CS) or IT (Polakowski, 2009). The

benefits of having a tertiary degree are echoed by Moore (2010) who suggests that having such a degree can be advantageous when looking to apply for positions requiring mid-level and upwards skills. Database administration is considered an advanced subject covered in universities (Simmonds, 2013). However, Simmonds also found that the confines of the lecture room posed some challenges when it related to providing students with practical exposure to the database environment.

2.3.4 Training methods

There is no one standard template used by DBAs to get the necessary skills to work as a DBA; they incorporate a variety of methods and materials into their training. Training includes courses that are offered by formal institutions such as universities and colleges as well as specialised training provided by the various database vendors currently present in the market (Peshkar & Ghosekar, 2015). Training by vendors such as Oracle, IBM, and Microsoft can provide introductory as well as advanced and specialised training courses, providing the DBA with the required technical skills to manage their database software. These are not only once off training courses but can be used continuously to stay abreast of technology changes (Peshkar & Ghosekar, 2015). In addition to the training courses are certification exams which can be written to get vendor product certification (Polakowski, 2009).

2.4 The big data environment

2.4.1 What is big data?

Big data is not a new phenomenon; the size of data has been a growing challenge since the 1970s, but there has been no set definition of the word 'big' in the term data (Borkar *et al.*, 2012a). The term "big data" first appeared in the early 1990's, previously, big data was defined only as the volume of structured data that could not be processed efficiently (Kaisler *et al.*, 2013). The term "big data" was later given in recognition for the growth and importance attributed to unstructured data that was not easily accommodated in traditional database systems and other reporting and analysis tools (Manyika *et al.*, 2011)**.** Hence, storage, analysis and processing of such data started to provide challenges to traditional database systems (Hashem *et al.*, 2015). Almeida and Calistru (2013:12) define the term big data "as a way of describing the phenomenon of growth in data volume, complexity and disparity". Big data is not only unstructured data but results from the explosion in growth of both structured and unstructured data from a diversity of sources (Hashem *et al.*, 2015). In essence, it is thought that big data is data that cannot properly be accommodated or processed by traditional database systems such as the RDBMS (Debortoli, Müller

& vom Brocke, 2014). Many organisations have thus started to realise the potential value that big data can bring to their organisation and are aiming to put plans in place to try to extract this value (Provost & Fawcett, 2013).

2.4.2 Composition of big data

Big data is divided into structured, semi-structured and unstructured data. Structured data resides in traditional RDBMSs in the form of rows and columns, while unstructured data presents a challenge for relational databases due to its varying nature (Courtney, 2012). Sources of big data include amongst others, weblogs from e-commerce websites, GPS coordinates from mobile devices, cameras, microphones, wireless sensor networks, content produced on social networks, emails, photographs, and audio files (Dijcks, 2011). Manyika *et al.* (2011) refer to data being produced as by-products from activities within organisations as they go about their daily business. Billions of people around the world are continually contributing to social media via their workstations, laptops, or smartphone devices. Already in 2012, big data consisted of less than 10% relational or structured data, and that percentage is on the decline with the continued growth in unstructured data (Bizer, Boncz, Brodie & Erling, 2012). The introduction of Web 2.0 and beyond has resulted in vast amounts of data being generated through social capital, incorporating data from the diverse social interactions of humans (Provost & Fawcett, 2013). Mishra and Sharma (2015) described big data as data generated from two distinct sources namely, humans and technology.

2.4.3 The growth of big data

The advanced manner in which data is being recorded and stored has resulted in this explosion in the quantity of data available (Diebold, 2003). The rate at which data is being generated has increased to such an extent that the internet is generating more data every second than that which it stored twenty years ago (McAfee & Brynjolfsson, 2012). By 2012, data warehouses had grown considerably, and some of the bigger data warehouses hold many petabytes of data and spread over many parallel databases, as is the case at eBay (Borkar *et al.*, 2012b).

2.4.4 Characteristics (the five Vs) of big data

Big data is characterised by volume, velocity, variety, value, and veracity (Sowmya & Sravanthi, 2017).

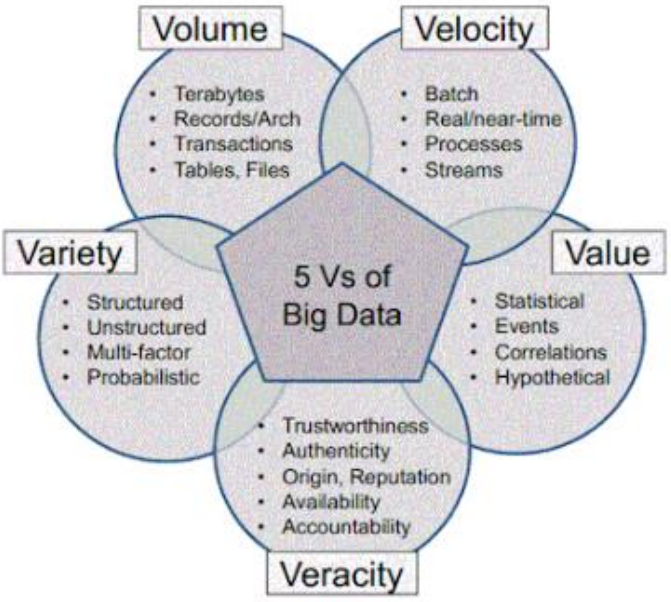


Figure 2-1: Characteristics of big data
(Sowmya & Sravanthi, 2017)

Volume

Machine generated data is continuously being produced at a much greater rate than data in traditional RDBMSs. Dijcks (2011) notes the example of a single jet engine having the ability to generate ten terabytes of data within 30 minutes. The continuous exchange of information through email messages, posts on social networking sites, information from mobile phones, audio files, web traffic all contribute to the growing amount of data which is currently available (Courtney, 2012).

Velocity

Velocity refers to the rate with which data is being produced. Millions of smartphone devices are storing their time and location every 15 minutes (Jacobs, 2009). While Twitter messages are only 140 characters in length, in 2011 it was noted that more than eight terabytes of data are stored every day due to the high frequency with which these messages are published (Dijcks, 2011).

Variety

Courtney (2012) emphasises that it is not the volume of the data that presents the biggest problem to organisations, but it is the variety and speed with which the data continually changes. There are numerous data formats being input into big data from a diversity of sources, amongst others these formats include text, videos and logs (Hashem *et al.*, 2015). Hitzler and Janowicz (2013) note that coupling this variety with the volume presented by big data, is what makes it challenging.

Value

Organisations need to filter and analyse this big volume of data in order to derive value. It will require the implementation of new technologies and analyses techniques in order to capture its value, as legacy systems and incompatible formats prevent these data types from being integrated with traditional enterprise data (Almeida & Calistru, 2013). In light of this, Storey and Song (2017:51) note that "[value] is the most time-consuming to extract, and difficult to ascertain".

Veracity

Data comes from many sources and in many different forms. As such, there is no standard with which the data arrives in the system regarding its cleanliness. Sowmya and Sravanthi (2017) refer to it as biases or noise. It is not always possible to ascertain the veracity of big data, due to the multitude of possible sources, and yet it is important to do so to make good business decisions.

2.4.5 The second economy

Even though big data provides numerous opportunities one concern noted by Botta, De Donato, Persico and Pescapé (2014) relates to the usefulness of the data being produced as the amount of data being produced is currently outweighing the ability to process and make use of such data. At a national level, defining a country's strength will be determined by its ability to exploit the potential of big data and not solely on its political and economic welfare (Mishra & Sharma, 2015). Organisations no longer treat big data as a by-product, but as an asset within the organisation and make special provision for it through infrastructure upgrades (Storey & Song, 2017).

2.4.6 Suitability of existing database systems

Tauro, Aravindh and Shreeharsha (2012) has also summarised the difficulties facing the RDBMS as being data growth over the last few years, increasing connectivity between data being produced on the internet and the nature of data being generated. Similarly Nayak, Poriya and Poojary (2013) state that the RDBMS is challenged when required to scale at the levels required by big data and leads to performance degrading over time as data volumes increase. The demands of web-scale systems requiring database systems to be able to scale on-demand and have the needed flexibility while still providing efficient performance, has posed a challenge for the RDBMS (Madison, Barnhill, Napier & Godin, 2015). Latency is a consideration for the RDBMS when required to scale by adding additional nodes as well as the added complexity of being able to store and analyse this volume of data

which can scale terabytes per day (Corbellini, Mateos, Zunino, Godoy & Schiaffino, 2017).

The increased volumes and new varieties of data types that databases need to accommodate have required database systems to evolve (Holt *et al.*, 2015). The exponential growth of data and the variety of data produced by various devices have posed a challenge for traditional relational databases hence the emergence of newer technologies to accommodate these requirements (Corbellini *et al.*, 2017). Traditional database architectures have a pre-defined structure into which the data needs to fit. Conversely, big data can be seen to work in reverse order whereby the structure is only created once the data has been loaded into the appropriate storage system and meta-data has been applied (Costa & Santos, 2017).

Costa and Santos (2017) outline a process undertaken by data in a big data environment as follows: (1) gathering data from the diversity of data sources, (2) loading data into the appropriate distributed storage architecture, (3) process and transform the data into meaningful information and (4) finally the data is consumed either through analytics, reporting and data extracts.

2.5 Big data architectures and implementations

Many different vendors supporting big data have brought about their own implementations of the Hadoop architecture and tools to provide a solution to the organisations' needs to extract value from data. Erraissi, Belangour and Tragha (2017) compare five distributors of Hadoop, namely HortonWorks, Cloudera, MapR, IBM, and Pivotal HD. They found that "most [providers] have created their own distributions relying on Apache Hadoop and associated open source projects" (Erraissi *et al.,* 2017:58). There is no one-size-fits-all implementation of a big data solution because of the different technologies available, making the landscape a tricky one to traverse. Organisations thus face several decisions when implementing a big data solution. Different technologies and architectures exist to meet different requirements, and it is often tricky for organisations to firstly understand and decide which technologies they require, and furthermore what skill set they require within individuals managing these environments (Volk *et al.*, 2017).

One of the big data architectures referenced by Costa and Santos (2017) is the NIST Big Data Reference Architecture (NBDRA). This conceptual model was designed by the NIST Big Data Public Working Group (NBD-PWG) Reference Architecture Subgroup in collaboration with other NBD-PWG subgroups (NBD-PWG, 2015). This

model was conceptualised from various big data architectures currently available (see Figure 2-2). The model intends to serve as a guide and framework to different business environments, to show the different components within a big data system in complementing current systems, without supporting or promoting any vendor or solution (NBD-PWG, 2015). The NDBRA model illustrates the complexity of the big data environment in relation to the many different technologies and tools needed to implement a big data solution.

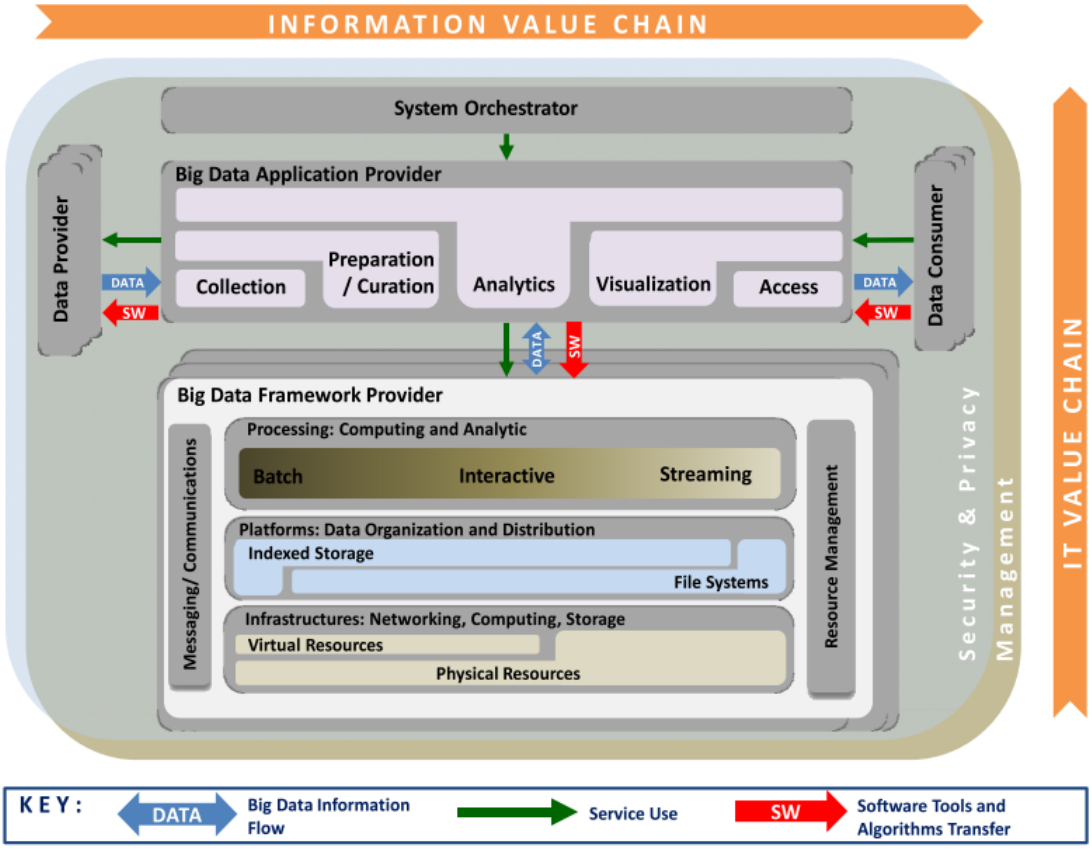


Figure 2-2: NIST Big Data Reference Architecture (NDBRA)
(NBD-PWG, 2015)

2.6 Big data technologies

Big data does not consist of a single set of tools but rather a comprise across a wide variety of technologies (Sharma, 2013). The technologies within big data can be broadly classified according to their usage namely: computing tools (e.g. distributed processing tools such as Hadoop); storage tools (e.g. non-relational databases such as NoSQL); and supporting tools (e.g. query languages to access the data such as SQL) (Almeida, 2017). There are multiple vendors within each category, each supporting a version of the technology with a different set of features or a different implementation of the standard features. Big data can also be classified according to different strategies required for processing the data (Sharma, 2013).

Below is just a brief introduction to the many tools and technologies that have arisen in the big data space in the recent past. It is not the intention to cover the entire big data technology stack but only provide a glimpse into the emergence of the various technologies and tools used to manage and accommodate big data. Of importance is the requirement for new technologies to enable the distributed processing, data storage and management of unstructured data within the big data environment.

2.6.1 Big data database solutions

Database systems have become an increasingly important part of an organisation's business. The focus has thus shifted towards the analytic capabilities which database systems need to include rather than only focusing on the transaction processing capabilities to enable enterprises to analyse customer buying patterns and trends (Arora & Gupta, 2012). The advent of social media platforms and other technological advances has also meant that databases are being used differently because of different types of data being shared among people (Holt *et al.*, 2015). Hence a new kind of database to meet the demands of big data is needed.

NoSQL databases

NoSQL ("Not Only SQL" or "Not Relational") database technologies have made the entrance onto the database software market in 1998 and have been developing since then (Madison *et al.*, 2015). NoSQL is one of the popular non-relational database technologies to emerge (Erraissi *et al.*, 2017). These database technologies have become available because of the inability of RDBMSs to effectively store unstructured data and manage unstructured data within the big data environment (Arora & Gupta, 2012; Priyanka & AmitPal, 2016). Unlike RDBMSs, NoSQL solutions do not categorise or parse the data, and increasingly accommodate the variety of data types available within big data. As with RDBMS whereby a schema is required to be in place when data is added, known as "schema-on-write", NoSQL databases are designed not to have this requirement and are schema-less, also known as "schema-on-read" (Storey & Song, 2017:56). NoSQL database technology avoids the expense of making use of "joins" at the data level and places the onus upon the application to perform this operation (Corbellini *et al.*, 2017). In order to reduce the complexity of making sense of the data, frameworks such as MapReduce have been introduced. In order to access the data, many different vendors have developed customised query languages to accommodate the requirements of their implementation of NoSQL (Nayak *et al.*, 2013).

The emergence of NoSQL databases will not signal the end of the RDBMS. However, in order for the enterprise to gain benefit from these NoSQL databases, there is a need for it to be combined with SQL solutions (Dijcks, 2011; Madison *et al.*, 2015).

ACID vs BASE guarantees

Some implementations of NoSQL databases have opted to conform to BASE guarantees rather than ACID guarantees (see 2.2.2) which have been associated with the RDBMS (Madison *et al.*, 2015). BASE stands for “basically available”, “soft state”, “eventually consistent”. “Basically available” improves availability of data through “replication and sharding”, “soft state” does not guarantee that data will be consistent but data across the system can change, and “eventual consistency” whereby changes made will eventually be consistent across the system (Binani *et al.*, 2016:44; Priyanka & AmitPal, 2016). “The idea behind the systems implementing this concept is to allow partial failures instead of a full system failure, which leads to a perception of a greater system availability” (Corbellini *et al.*, 2017:6). The requirement for the RDBMS to comply with ACID guarantees places a strain on its ability to scale and process data in a distributed fashion as is needed within a big data environment. Replication of data is important as nodes often fail within such distributed systems, and storage systems with improved fault tolerance, reliability and scalability are now required to meet some of the challenges posed by big data (Siddiqa, Karim & Gani, 2017).

In order to accommodate increasing volumes of data, systems can cluster several nodes together to allow it to be perceived as a single node and this is termed clustering (Priyanka & AmitPal, 2016). However, clustering has a concern regarding consistency, and hence NoSQL databases have adopted the Brewer’s CAP theorem which stands for “consistency”, “availability” and “partition tolerance” (Nayak *et al.*, 2013:16; Priyanka & AmitPal, 2016). “Consistency” relates to all nodes being in the same state, “availability” to retrieve data even though it might not be consistent after a failure and “partitioning tolerance” to enable writing of data to multiple nodes (Priyanka & AmitPal, 2016:4963-4964).

Big data storage systems are challenged between the requirements of data being consistent versus data being highly available, but the partitioning element of the theorem is mandatory. Such systems will be classified either as being available and partition-resilient (AP), or consistent and partition-resilient (CP). Thus, systems conforming to AP place a higher priority on replication to ensure availability of data,

while systems conforming to CP place a higher priority on ensuring data is consistent across the nodes (Siddiqa *et al.*, 2017).

Horizontal vs vertical scalability

NoSQL databases are horizontally scalable and non-relational, having the ability to store structured data like an RDBMS while including semi-structured and unstructured data (Arora & Gupta, 2012). Horizontal scaling enables these databases to scale-out by adding commodity hardware rather than scaling up by replacing existing with bigger and more powerful hardware (Storey & Song, 2017). This provides many advantages such as lower cost, enhanced scalability, load balancing, and improved processing power through parallelisation. The ever-decreasing cost and improvements to hardware components including storage, CPU and memory as compared to a couple of decades ago has also helped in this regard (Storey & Song, 2017).

Scalability does not just refer to the expansion of the physical hardware but also includes the ability to accommodate a variety of data (Lee *et al.*, 2013; Priyanka & AmitPal, 2016). Increased load caused by more concurrent users accessing the systems is thus not constrained in a central place but is spread across the distributed architecture (Zaki, 2014). This is achieved through mechanisms known as sharding whereby data is split into smaller parts or through replication which mirrors the data to improve availability at the cost of storage (Corbellini *et al.*, 2017).

2.6.2 Categories of NoSQL databases

There are many variations of NoSQL databases each being designed towards a specific function and problem it is trying to solve (Raut, 2017). They can generally be categorised in four groups namely key-value stores, column-orientated databases, document-based stores and graph databases with each having a different approach (Moniruzzaman & Hossain, 2013; Storey & Song, 2017).

Key-value stores

In key-value stores, databases store the data based on keys which are indexed (see Figure 2-3). These keys are then used to retrieve the data (Leavitt, 2010) while not making use of schemas as in the case of the RDBMS (Nayak *et al.*, 2013). In its simplest form, the key is a pointer to the data or value to be accessed, and both key and value are stored as a Blob (Binary Large Object). The data within these databases is accessed through what is known as ‘get’ and ‘put’ operations (Pokorný, 2013) and the use of SQL is not excluded (Storey & Song, 2017). Examples include

Voldemart developed by LinkedIn, Berkeley DB developed by Oracle, Tokyo Tyrant and many more.

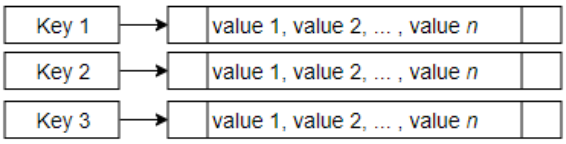


Figure 2-3: Key-value store

Column stores

Column store databases (see Figure 2-4) use a single column instead of rows which contain similarly related data (Leavitt, 2010). Functions of these databases include the ability to version-control changes to data as well as the support of predictive analytics (Madison *et al.*, 2015) and the processing of large volumes of data (Raut, 2017). "[It] is an extension of the Key-Value Store database where columns can have a complex structure, rather than a blob value" (Storey & Song, 2017:58). An example of such a database is the Cassandra database created by Facebook.

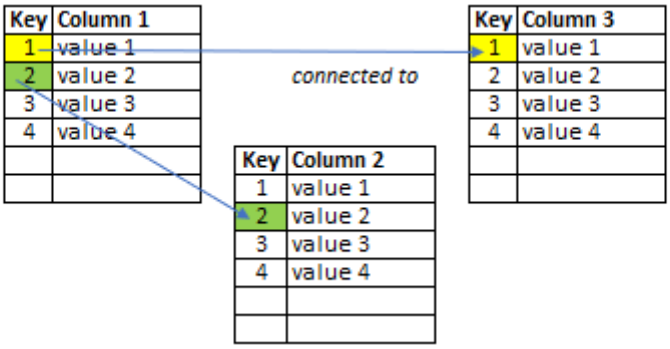


Figure 2-4: Column store

Document stores

The primary goal of document store databases (see Figure 2-5) is to store data as documents (Moniruzzaman & Hossain, 2013) - the concept of documents and collections being likened to records and tables respectively within an RDBMS (Raut, 2017). These databases make use of a key which points to a document and can support structured, semi-structured as well as unstructured data. Data types supported by these databases include XML (eXtensible Markup Language), JSON (JavaScript Object Notation) and BSON (Binary JSON) (Corbellini *et al.*, 2017; Raut, 2017). Examples of these types of NoSQL databases are CouchDB and MongoDB (Priyanka & AmitPal, 2016).

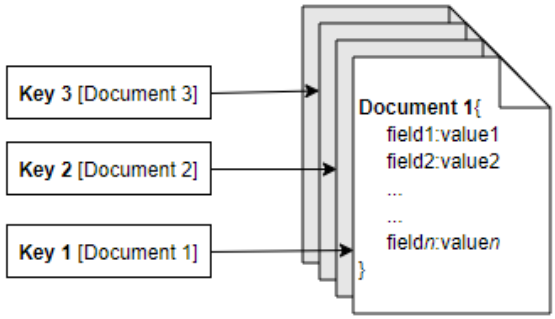


Figure 2-5: Document store

Graph databases

Graph databases (see Figure 2-6) are described as non-relational databases, with the addition of also supporting ACID guarantees like relational databases (Storey & Song, 2017). Data within these databases represent the data as 'nodes', while 'entities' are the relationship between those nodes as well as 'properties' which is information relevant to the node (Priyanka & AmitPal, 2016). One of the concerns with this database is its ability to scale-on-demand at runtime (Siddiqa *et al.*, 2017). The concept behind these databases is to show networks of connections and examples are Neo4J and HyperGraphDB (Raut, 2017).

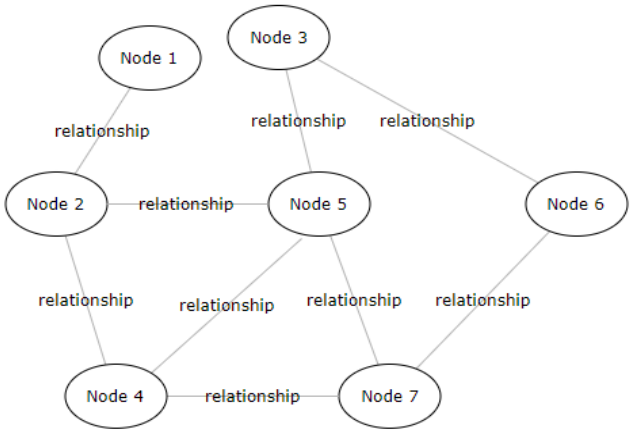


Figure 2-6: Graph database

Advantages of NoSQL databases

The increasing emergence of distributed systems via web applications such as Facebook, Amazon and many others has placed a different requirement on the databases servicing these systems. The rigid enforcement and support of ACID guarantees by the RDBMS make it inefficient for it to support these distributed systems requiring high availability and performance (Hecht & Jablonski, 2011; Storey & Song, 2017). The unstructured nature of big data has put a new requirement on databases to become more scalable and more flexible, and this is a short-coming of RDBMS where data needs to be modelled within the limitations of a static schema

(Zaki, 2014). In comparison to the RDBMS, NoSQL databases have made it easier to scale out through low-cost commodity hardware, ability to better accommodate the requirements of big data and while no reliance on expensive proprietary infrastructure (Raut, 2017).

Limitations of NoSQL databases

According to Atzeni *et al.* (2013), these databases are more difficult to maintain as there is no separation within the database into logical and physical structures. The NoSQL databases are quite complexed to set up and maintain in addition to requiring object-oriented programming knowledge rather than knowledge of databases (Madison *et al.*, 2015). There are a variety of implementations and no standard method of querying across the different variations (Corbellini *et al.*, 2017). Some notable differences that users have to become accustomed to include the de-normalised state of the database, and the absence of joins and foreign keys (Storey & Song, 2017). This flexibility does provide an advantage over the RDBMS because of its loosely defined structure; there is very little if any impact on the end-users when making any change to a relational model.

Another concern of NoSQL databases is limited security, as these databases focus on scalability and performance, and the onus is put on developers or users of these databases to implement security within other tiers of the system (Zaki, 2014).

Despite all the features and abilities that NoSQL provides to enterprises, Kumar and Charu (2014) note that the lack of adherence to ACID guarantees is a potential obstacle in preventing large enterprises from adopting NoSQL as an alternative to the traditional RDBMS. Storey and Song (2017) agree with the lack of support for ACID guarantees in addition to the fact that it is not a platform that supports the requirements of an OLTP system. New implementations of NoSQL databases are resolving these challenges and are being developed with the needed support for ACID properties. The main focus is thus the ability to process vast amounts of data while sacrificing data consistency through the property of eventual consistency (Pokorný, 2013).

NoSQL databases are still evolving, and certain requirements are currently being fulfilled and will continue to be fulfilled by relational databases hence it is not one or the other but both these technologies co-existing alongside one another (Storey & Song, 2017).

2.6.3 NewSQL databases

NewSQL database systems have been born from a combination of properties supported by both RDBMS and NoSQL systems. These include a combination of many features available in RDBMS and NoSQL systems such as the support of ACID guarantees available in the RDBMS and scalability through a distributed architecture supported by NoSQL systems (Silva *et al.*, 2016; Storey & Song, 2017). NewSQL also incorporates the vertical and horizontal scaling capabilities of RDBMS and NoSQL database technologies respectively (Binani *et al.*, 2016). While many vendors have implemented their own language to query these databases, many have also incorporated support for SQL as a query language, for example VoltDB (Silva *et al.*, 2016).

Vendors have opted to follow various approaches to NewSQL databases while still maintaining the requirements for both big data and traditional RDBMS (Kumar & Charu, 2014). NewSQL databases are classified according to the following groups: new architectures; optimised SQL engines and; transparent sharding (Kumar & Charu, 2014; Silva *et al.*, 2016). “New architecture databases” make use of fast storage devices like in-memory, flash and SSD. “Optimised SQL engines” provide new engines with better scalability while “transparent sharding” enables databases to scale by spreading them across multiple nodes (Kumar & Charu, 2014; Silva *et al.*, 2016:417). An example of NewSQL is Google Spanner (Binani *et al.*, 2016).

2.6.4 Massively parallel processing

The storage of unstructured data now requires databases with a columnar architecture unlike that of RDBMSs (Sharma, 2013). These large volumes of data are processed by being broken down into smaller chunks and processed in parallel across tens, hundreds and even thousands of servers using the strategy known as massively parallel processing (MPP) (Jacobs, 2009). This uses the “shared nothing” architecture in order to perform parallel processing across many nodes but Arnold, Glavic and Raicu (2015) note that in general MPP only works for implementations with a relatively small number of nodes in the cluster, hence questioning its ability to scale effectively to accommodate the increasing volumes of data being processed.

2.6.5 Hadoop

Hadoop is an open-source framework for distributed processing of large volumes of data developed by Apache Software Foundation (Hashem *et al.*, 2015), and it is an important technology when implementing a big data solution (Volk *et al.*, 2017). It consists of two core components namely the MapReduce programming framework

and Hadoop Distributed File System (HDFS) (Hashem *et al.*, 2015; Alguliyev, Gasimova & Abbasli, 2017). Data is processed in batches and across a cluster of commodity servers, each having its own memory and disk storage. There is thus no sharing of resources resulting in data being stored in multiple pieces across the various servers (Purcell, 2014). The advantage that it presents is scalability, load balancing and the ability to process large volumes of data in parallel (Storey & Song, 2017). Despite these advantages Hadoop has not been as widely adopted as initially expected, with in-memory platforms such as Spark gaining popularity at a faster pace (Storey & Song, 2017). In order to ease the management of Hadoop, vendors also incorporate management consoles to assist in this regard (Erraissi *et al.*, 2017).

Hadoop distributed file system (HDFS)

The Hadoop implementation also includes the Hadoop Distributed File System (HDFS) (Pingle, Kohli, Kamat & Poladia, 2012; Hashem *et al.*, 2015). HDFS is a distributed file system for storing large datasets with the ability to run on low-cost commodity hardware, scale across thousands of servers and is highly fault tolerant (Pingle *et al.*, 2012; Hashem *et al.*, 2015). As such, one of the characteristics of these systems is its ability to auto-detect and correct faults in an expeditious manner (Siddiqa *et al.*, 2017).

HDFS achieves distributed processing by dividing files into smaller blocks and distributing it across multiple nodes within the cluster (Costa & Santos, 2017). For example, HDFS is born from the Google File System (GFS) (Siddiqa *et al.*, 2017) and in GFS data is distributed across the system in 64MB chunks with an associated checksum (Liu & Dong, 2012). The architecture consists of a master and slave architecture, with large datasets being processed in smaller quotas across the slave nodes in the cluster (Gupta & Siddiqui, 2013). The master node which is the central coordinator or controller of the cluster is also referred to as the ‘namenode’ and is responsible for file access control. The slave nodes or ‘datanodes’ handle the data storage on the node itself.

MapReduce

MapReduce is a commonly used processing framework designed by Google for processing data in parallel on large clusters of commodity machines which can span to tens of thousands of machines (Hashem *et al.*, 2015; Silva *et al.*, 2016). MapReduce works alongside HDFS, whereby the blocks distributed across the nodes within the HDFS cluster is processed by the MapReduce framework (Costa & Santos, 2017). The data is firstly mapped across the different nodes for parallel

processing using key-value pairs and thereafter using a function called “reduce” which combines the result sets for the same key from the various nodes into a single result (Hashem *et al.*, 2015).

MapReduce is available from many vendors as either an open-source or commercial implementation with Hadoop being the most popular open-source version (Stonebraker *et al.*, 2010). Many open-source and traditional IT vendors now combine their database technologies with the Hadoop distribution to provide a strategy to deal with the complexities of big data (Sharma, 2013). Stonebraker *et al.* (2010) suggest that rather than MapReduce being a competitor of parallel database systems, it is best suited to work alongside these database systems as it is more like an extract-transform-load (ETL) system.

Other Hadoop technologies

Technologies within the big data landscape is continuously evolving and will continue to evolve. One of the key features introduced within Hadoop 2.0 is YARN (Yet Another Resource Negotiator) (Abadi *et al.*, 2016). The primary function of YARN is the management of computational resources within a cluster (Nizam & Hassan, 2017). It has brought about several enhancements over the existing MapReduce framework, with improvements in the areas of scalability, parallelism and resource management (Oussous, Benjelloun, Lahcen & Belfkih, 2018). It consists of components such as a global ResourceManager to manage the cluster, a NodeManager for each node and an ApplicationMaster for each application to be executed (Oussous *et al.*, 2018).

There are also other components within the Hadoop ecosystem such as Apache Mahout, Apache Sqoop and Apache Hive amongst others. Mahout is a distributed and scalable machine learning and data mining library which can executed by MapReduce (Madden, 2012; Khan *et al.*, 2014). Mahout consists of various algorithms for the purpose of “clustering, classification, and batch-based collaborative filtering” (Hu, Wen, Chua & Li, 2014:679). Sqoop is a data integration tool that supports the bulk transfer of data between Hadoop data storage systems such as NoSQL databases, and structured data storage systems such as relational databases (Hu *et al.*, 2014; Sowmya & Sravanthi, 2017). It is presented as a command-line interface (CLI) while also being known for its “fast performance, fault tolerance and optimal system utilization” (Oussous *et al.*, 2018:439). Hive, developed by Facebook, is a data management or data warehousing tool within the Hadoop platform used for data mining. It has its own query language to access the data,

HiveQL or HQL, and it is similar to SQL (Sowmya & Sravanthi, 2017). Its primary function is to support the analysis, summarization and querying of large distributed datasets and executes via MapReduce (Hu *et al.*, 2014).

2.7 Cloud computing

The emergence of cloud-based architecture has provided a platform and technology stack to enable the complex computations and analysis of large datasets (Hashem *et al.*, 2015). This new and flexible model of provisioning systems enables organisations the ability to scale-on-demand, achieve quick provisioning time and avoid large capital outlays for hardware and software. It also provides extensive scalability in terms of processing power and storage (Hashem *et al.*, 2015). This is accomplished through the costing model implemented by cloud providers whereby cost is determined by usage (Storey & Song, 2017). The lower cost is also enabled through cloud's primary adoption of virtualisation, enabling the cloud service providers to provide computing resources as and when required by the end-user (Zanoon, Al-Haj & Khwaldeh, 2017). This realises a lower cost of ownership considering all the costs relating to owning one's own hardware and software.

Cloud-based environments can either be based on physical or virtual machines. Users access environments and services through web-based front-ends or mobile devices (Zaki, 2014). These servers are situated in many remote locations upon which users theoretically have extensive resource capacity at their disposal and the ability to customise to their requirements (Mai, Nurminen & Di Francesco, 2014).

Cloud computing is a common component within a big data architecture as it provides the infrastructure upon which the solution is built (Volk *et al.*, 2017). The distributed nature of cloud-based environments whereby resources are geographically dispersed lends itself to the requirements of big data because data is being generated from everywhere (Zanoon *et al.*, 2017). Many of the requirements of big data are achieved through virtualisation technologies adopted by cloud computing. "Virtualization is a process of resource sharing and isolation of underlying hardware to increase computer resource utilisation, efficiency, and scalability" (Hashem *et al.*, 2015:99). It simulates a computing environment with operating systems and applications, much like what would be experienced should it have been a physical server (Zanoon *et al.*, 2017).

There are different types of clouds platforms available as well as services being offered through the cloud infrastructure, including private, public and hybrid clouds

(Purcell, 2014). Cloud services provided on these platforms include Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS). PaaS supplies the hardware and software to enable development. SaaS supplies pre-installed software and applications that are ready for end-users to use via the internet. IaaS provides the hardware infrastructure that end-users can access (Hashem *et al.*, 2015). These service models are supported by a variety of cloud vendors such as Google, Amazon, and Microsoft to name a few.

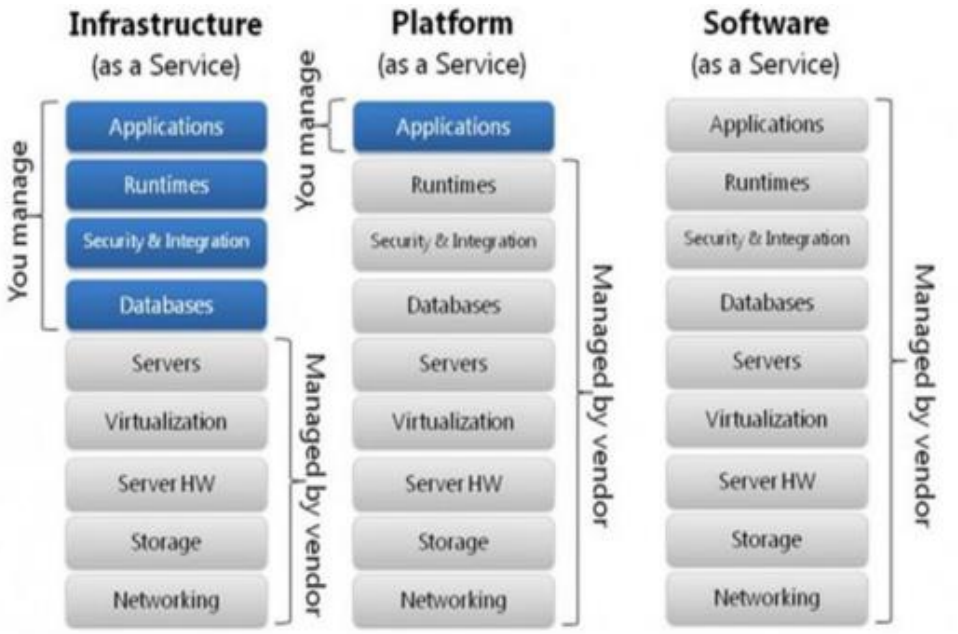


Figure 2-7: Cloud services
(Ahmed, 2015)

Capabilities of cloud computing

The flexibility and scalability of the cloud enable the storage of structured data used by the traditional RDBMS while providing the capabilities of also accommodating both semi-structured and unstructured data of which big data is comprised (Arora & Gupta, 2012). Increased attention is being given to cloud-based environments for its storage and processing capabilities, to handle the distributed storage and processing requirements of big data analytics. These environments can either be based on open-source solutions such as Apache Hadoop or propriety solutions from vendors such as Microsoft, Amazon, IBM, Google, and many others (Talia, 2013; Hashem *et al.*, 2015). McKinsey (cited in Zaslavsky, Perera & Georgakopoulos, 2013) reports that the big difference with cloud-based platforms is the ability to scale on demand using distributed file systems and parallel-processing databases, unlike the limitations regarding scalability and the cost thereof of traditional database environments. Usage of cloud-based environments has also been where

organisations used these platforms to test and develop new initiatives before making costly hardware and software acquisitions (Purcell, 2014).

Concerns with cloud computing

There are still factors which deter organisations from adopting cloud-based solutions, such as the potential for malicious activities to be conducted from such environments (Purcell, 2014). Holt *et al.* (2015) found that organisations still had concerns about vendor trust regarding the reliability and security of their data and the fact that it does not completely satisfy their requirements. Even as recently as 2017, issues of security are still a concern regarding the use of cloud computing (Storey & Song, 2017).

Database as a service (DBaaS)

The addition of databases in the cloud, known as Database-as-a-Service (DBaaS), has enabled consumers to store their data remotely through the internet, without the need for large capital investment on hardware infrastructure to house their databases. In addition to the reduction in investment, cloud databases offer a high degree of scalability, availability, and optimised resource usage through the use of a "pay-as-you-go" model (Arora & Gupta, 2012:77). This service provided by cloud service providers takes care of the setup and maintenance of the database (Corbellini *et al.*, 2017). The introduction of DBaaS has seen more attention being given to NoSQL databases (Nayak *et al.*, 2013). Many cloud services make use of NoSQL databases and DBaaS as their storage systems (Corbellini *et al.*, 2017).

2.8 Data science

Traditionally organisations have managed large volumes of structured data and relied on statistics to analyse and generate useful information in order to gain insight into, and make decisions about, their future business (Aasheim, Williams, Rutner & Gardiner, 2015). With the emergence of big data, organisations are now faced with the challenge posed to them by the unstructured nature of this data. The challenge is emphasised by the characteristics of big data concerning the five "Vs", and it is thus difficult to manage and analyse this data through traditional means. Organisations now need new technologies implemented within the big data environment and accompanying analytic methods to gain the insights they require to drive business value and new product lines.

Data science is considered the area which is primarily focused and responsible for deriving meaning from the unstructured data within an organisation, using the

various technologies within the big data environment (Costa & Santos, 2017). There is thus a relationship between data science and big data, whereby the big data technologies support and drive the vast analytics of data to useful information (Provost & Fawcett, 2013). Cao (2016:72) defines data science as "the science (or study) of data" which enables the creation of valuable insights and information and thereby drives potentially new organisational business value from the voluminous amounts of data within the organisation.

A key component of data science is data analytics, which consists of various disciplines such as machine learning, data mining and statistics (Aasheim *et al.*, 2015). Data analytic thinking is an important component of data science which in turn is an important component that needs to be embraced by the organisation in order to drive the competitive advantage that organisations seek to explore (Provost & Fawcett, 2013). In a traditional RDBMS environment, there is a single team of individuals which is responsible for the governance of the organisation's data. Within a big data environment, that responsibility is spread across various functional units within the organisation, each providing a specific set of skills and insight into interpreting the data so that it can become valuable information to the organisation. There is thus a close collaboration between the business and the IT units of the organisation (Vemuganti, 2013). As a result, data analytics does not only live within the realm of the data science team but also incorporates the various business units within the organisation that must understand the importance of data analytics to extract business value (Provost & Fawcett, 2013).

2.9 The data scientist

In the world of big data and data science, there is a crucial new role that has emerged, namely the 'Data Scientist'. The primary concern of a data scientist is the analysis of the data to drive competitive advantage which the business hopes to achieve through the transformation of the diverse data into useful information (Provost & Fawcett, 2013). They are responsible for analysing the organisation's data, both structured and unstructured, in an attempt to develop models and algorithms that allow the organisation to predict behaviours and patterns of their customers (Sukumar *et al.*, 2013).

The role of a data scientist is multidisciplinary, having many responsibilities and accompanying skills to perform those responsibilities. Almeida (2017:17) defines the data scientist as being "a professional with multidisciplinary skills in areas such as data engineering, scientific method, math, statistics, advanced computing,

visualisation, hacker mindset and domain expertise". Provost and Fawcett (2013) also list competencies such as data analytics, data mining, data processing as well as big data storage tools and technologies supporting data science. Thus, in many instances, the role of the data scientist has been associated with someone having a diversity of skills, an "all-rounder" (Murawski & Bick, 2017:1378). Costa and Santos (2017) noted that one of the many challenges that will be faced by organisations concerning the adoption of big data is the need for new skills to effectively manage big data. It is possible, given all the skills requirements of the data scientist role, that those skills are potentially in short supply.

2.10 Additional responsibilities imposed on DBAs

The emergence of big data has presented some challenges regarding the management of such a large and diverse dataset. It has been shown that there is a lack of knowledge throughout the entire organisation to effectively deal with the increase of big data (Bloem *et al.*, 2012). It has also had an impact on the functions performed and skills requirements of the traditional DBA (Simmonds, 2013). Additional job requirements now imposed on the DBA due to a shift towards big data has resulted in an overlap with those of systems administrators, data analysts, developers and network administrators, all because of the technology stack required to implement a big data solution (Simmonds, 2013). The rise of the data scientist has further complicated the skill set requirements of the DBA as it is a role with a multidisciplinary skill set within the big data environment.

2.11 Summary of Chapter Two

Chapter Two looked at the literature to understand the research problem. There are two main components reported in the literature, namely the RDBMS environment as well as the big data environment. In addition to the RDBMS, the role of the DBA was described in terms of their skill set, responsibilities, and training. The primary technology reviewed was the relational database which has been in existence for some time now.

It is evident that the volumes, variety, velocity, and veracity of data being produced are continuously increasing, and while the RDBMS has been in existence for a while, the RDBMS is facing challenges in meeting the requirements of processing big data. While in many instances the RDBMS can manage the volume and velocity of structured data, it is the complexity added through the increased requirement to extract value from unstructured data that has presented the RDBMS with its biggest challenge.

Big data is characterised by volume, variety, velocity, value, and veracity which consists of structured, semi-structured as well as unstructured data. These diverse data types have required a new set of technologies in order to manage and analyse it in order for it to be of value to organisations. As such, the big data environment consists of many different technologies and tools to manage and analyse this data, each having a specific function to derive value from the volumes of data. New technologies including distributed processing frameworks (e.g. Hadoop) and non-relational database technologies (e.g. NoSQL and NewSQL) have become prominent in recent years due to their ability to store and process these volumes of data in a distributed manner. These technologies also incorporate the flexibility and scalability required by big data, allowing them to scale on-demand using commodity hardware. In addition to these technologies, the emergence of the cloud has provided in many instances the infrastructure to accommodate the big data technology stack. A flexible platform with extensive resources using virtualisation has become a key feature within the technology stack to manage and process big data.

The intention of the literature review was not to describe in depth each technology, nor was it the intention to provide an in-depth technical assessment of the key technologies identified. The goal was to provide a high-level overview of the technologies within the big data environment which would in most instances be associated with functions to be performed by a DBA in order to identify where in this environment the DBA can occupy a similar role. Chapter Three will present the research design and methodology followed by this study.

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

Chapter Two presented an overview of the literature in relation to the RDBMS and big data environments. It included the roles of the DBA within the RDBMS environment as well as the emerging role of the data scientist within the big data environment. This chapter will provide insight as to the research process to be followed and the disciplines that underpin this research. Undertaking research is an attempt to understand or contribute towards a possible solution for a research problem, and therefore the problem identified is important in guiding the research (Ellis & Levy, 2009). The researcher undertakes to provide the reader with an understanding of what has driven the chosen philosophical paradigm, research methodology, research design as well as the data collection methods.

Creswell (cited in Tuli, 2010) noted that there are different factors which influence the chosen research methodology. These include the research problem, the researcher as well as research's audience. Research studies are also based upon influencing factors such as the research problem, literature supporting the problem and the availability of data to be collected (Ellis & Levy, 2009).

Many factors contribute to and influence the way in which researchers understand and experience the world in which they live. These can be through their own experiences and situations they find themselves in, cultural views, and even their own working experience can be different from others around them (Tuli, 2010). This is also no different for participants taking part in a research study. The different influences as experienced by researchers will have a similar impact on the way in which DBAs experience their working environment.

In determining the appropriate philosophical stance to be selected for this study, it is necessary to identify the context in which the study is to be conducted. In the background of this study, the need for a new job role was established to administer a big data environment, namely the data scientist (Davenport & Patil, 2012). They further state that little is currently understood as to whether this role will be fulfilled by a single individual with a variety of skills or many people and as a result, they pose the question, "What kind of person does all this?" Currently, the DBA is responsible for managing and administering the database within which the organisation's data exist. Many new technologies for storage, management, and processing big data have emerged within the big data environment (section 2.6) and as such it is unclear

as to where within this environment the DBA fits. Considering the above, it is unclear as to the impact that this new role will have on the DBA transitioning from the RDBMS environment to a similar role within a big data environment.

3.2 Research philosophy

In an attempt to address the research questions identified in this study, it is required to decide between two of the primary philosophical paradigms, namely subjectivist or objectivist. Holden and Lynch (2004:3) note that these paradigms are "delineated by several core assumptions concerning ontology (reality) and epistemology (knowledge)". A philosophical assumption is how the researcher sees or interprets that which occurs in the world, and this view is based on the researcher's ideologies (Remenyi *et al.*, 2011). However, to understand how to conduct research regarding which methodologies and methods to choose, it is imperative to understand why the research needs to be conducted (Holden & Lynch, 2004). The context of the study and questions to be answered influence the philosophical approach taken (Crossan, 2003). Remenyi *et al.* (2011) emphasise that it is of utmost importance for any researcher to understand the underlying philosophical assumptions that are the basis of any research activity. "Thus an innate understanding of philosophy is a cornerstone of knowledge creation" (Remenyi *et al.*, 2011:356).

In order to determine the appropriate research methodology to use in a study, Easterby-Smith (cited in Crossan, 2003) refers to three points as to the reason why a philosophical stance is important:

- It helps to identify the research strategy to be implemented regarding the type of data to be collected and the origin of such data.
- It quickly identifies the short-comings of various research strategies.
- It aids the researcher in exploring other methods outside of their expertise.

The importance of the researcher's epistemological perspective must not be taken for granted as it has an underlying influence on the theoretical perspective, methodology and methods chosen for the study (Gray, 2013:19). Positivist and interpretivist paradigms are influenced by their underlying ontological and epistemological beliefs (Walsham, 1995).

3.2.1 Ontology

Ontology is the view of what *is* reality (Walsham, 1995). Neumann (cited in Tuli, 2010) states that within the ontological paradigm, there are two broad and differing stances concerning the nature of reality, the first being "objectivism" and the second

being “constructionism” (also known as subjectivism or interpretivism). Objectivism dictates that the nature of reality is not constructed through the participation of human experiences, but it exists independently. Constructionism dictates that reality is influenced by human action. The constructionism ontological stance is aligned with that of the interpretivist epistemological stance in that there is a degree of participation from the human perspective regarding the nature of reality and knowledge of such reality, respectively. The nature of reality within the constructionism stance is therefore not independent and isolated from individuals, and often reality is constructed through the lens of the individual experiencing that reality.

Given the above, a constructionalist ontological stance has been adopted for this study.

3.2.2 Epistemology

Holden and Lynch (2004:7) state that “the relevant epistemological stance is that knowledge cannot be discovered, as it is subjectively acquired – everything is relative”. Burrell and Morgan (cited in Morgan & Smircich, 1980) depict the extreme subjective view to epistemology as being insight and understanding. The positivist’s approach to research is guided by the fact that it is objectively defined in that the researcher has no influence on and is removed from the phenomenon being observed in order to increase the predictability of the outcome when testing a theory (Myers, 1997; Crossan, 2003).

3.2.2.1 Positivist paradigm

Positivism dictates that reality is independent of the researcher and that it should be studied through scientific enquiry (Gray, 2013:20). Rather than attempting to explore and gain an in-depth understanding of the subject area, positivism approaches the study within a quantitative manner (Crossan, 2003). Knowledge is thus generated scientifically through numeric measures (Wahyuni, 2012). The firm belief that the outcome of research must be repeatable is at the heart of positivism (Wahyuni, 2012). Thus a degree of predictability is needed in this approach because of the requirement for empirical evidence as a source of knowledge. It is not possible to determine the level of engagement with big data of the DBAs across the various organisations within this study, and thus the level of experience that respondents have will be different. Also, technology is continuously evolving in the big data arena. Hence, the same questions posed to respondents over a long enough period would bring about change in their responses to those questions.

This research cannot easily adopt a positivist epistemological approach as the study's objective is not to determine causality against a set of fixed rules (Morgan & Smircich, 1980).

3.2.2.2 Interpretivist paradigm

The interpretivist epistemological stance is that social actors influence reality, and thus reality is the way these actors view it. As such, this reality is also influenced by the different experiences and backgrounds of such individuals (Wahyuni, 2012). Interpretivism prescribes that there is a difference between natural and social realities and they require different methods of enquiry (Gray, 2013:23). Walsham (1995) notes that within interpretive studies the researcher should be flexible in their initial approach based on the data collected within the study.

Interpretivists do not conform to a single truth and are of the view that reality is subject to change. Interaction with the subject is preferred in order to understand the subjective nature of reality from the subject's perspective rather than being separated from the subject as prescribed within the positivist approach (Wahyuni, 2012). Hence the researcher in this study is not making any assumptions as to the extent to which DBAs have received exposure to big data. Getting to understand what constitutes reality is more important than the ability to generalise the results (Gray, 2013:34).

Interpretivist researchers look at various situations unfolding in the real world, and this allows situations to unfold rather than scientifically controlling them (Tuli, 2010). As previously mentioned, different organisations will adopt big data technologies according to their specific requirements and this is outside of the control of DBAs. Their level of involvement, experience and ability cannot be determined or measured against a fixed set of rules. Their real-world experiences and expertise within the role itself are different, and thus each will adapt to the big data environment differently. An interpretivist stance allows for the outcome to be influenced by the context of the study and its participants to gain such an understanding.

With the above in mind, an interpretivist epistemological stance has been adopted.

3.3 Research approach

3.3.1 Qualitative research methodology

Morgan and Smircich (1980:498) refer to quantitative methods as being “appropriate for capturing a view of the social world as a concrete structure”. They go on to note that typical methods include large sample-sized surveys and laboratory experiments based on assumptions from an objectivist point of view. Quantitative and qualitative research methodologies each must satisfy several different aims in order to validate the outcome of the research. One of the aims of quantitative research is to describe, and this is done against a fixed set of constraints that are aimed at being repeatable and controlled. To the contrary, qualitative research aims to find deeper understanding through exploration (Tuli, 2010). However, many believe that qualitative research suffers from being influenced by researcher bias given that different researchers can potentially produce vastly different outcomes. Thus, the validity and reliability of qualitative research in many respects are questioned (Mays & Pope, 1995). Kaplan and Maxwell (2005) suggest that there is a balance between qualitative and quantitative studies in that validity is more prominent within qualitative studies while reliability is more evident within quantitative studies.

Qualitative research is often used within studies relating to information systems in respect of the people and the systems they use (Kaplan & Maxwell, 2005). Qualitative methods look at the social aspect of the entity being studied (Myers, 1997). Tuli (2010:103) further states that “qualitative researchers attempt to understand the complexities of the world through participants’ experiences”. The organisational context also influences the way in which information systems are used (Kaplan & Maxwell, 2005).

The qualitative approach assesses the outcome of the study through the interpretation of data as opposed to quantitative studies which rely on measurements (Stake, 2013). Qualitative research uses data collection methods that enable the researcher to delve deeper into the environment in which the participant finds themselves (Tuli, 2010). In this case, semi-structured interviews will be used.

This research will thus be conducted using a qualitative methodology.

3.3.2 Inductive approach

Two main research approaches can be adopted within a research study namely inductive or deductive. Inductive reasoning moves from a relatively unknown or

unfamiliar understanding of a concept or idea towards a better understanding. Through the collection and analyses of data the researcher can get a better understanding of the subject area to start moving towards theories and generalisations rather than to prove or disprove existing theories (Gray, 2013:17). Deductive reasoning, on the other hand, begins with an existing view or proof and the intention is to test or confirm a hypothesis (Gray, 2013:16). Deductive reasoning is associated with concrete, observable and empirical evidence with the intention to accept or reject the hypothesis and does not rely on subjective evidence (Gray, 2013:16). As such, a large proportion of quantitative research is considered to be deductive (Atieno, 2009), while qualitative research is primarily inductive and allows theories, models and frameworks to emerge from the data (Kaplan & Maxwell, 2005).

In light of this study being qualitative and with no initial hypothesis or theory to prove or falsify, it is the intention of the research thus to consider an inductive approach to this research.

3.4 Research design

Following the qualitative research methodology this study will follow, exploratory research will be conducted. It is important to note that exploratory research is not strictly nor only associated with qualitative research but can also be used within quantitative studies (Jupp, 2006).

3.4.1 Exploratory research

An important aspect of research is to identify the purpose thereof. Exploratory studies are aimed at finding more information about a relatively unknown subject area in order to evaluate the viability of further exploration (Gray, 2013:36). Exploratory research allows the researcher some degree of freedom in this context, but it is also important to note that such research should not be conducted in a haphazard manner (Jupp, 2006).

Ellis and Levy (2009) suggest that research can be categorised into three different types namely; research that is looking to build towards a theory, research attempting to validate theories and research looking to revise existing theories. Exploratory research thus aims towards building theories from the data that emerges from the study (Jupp, 2006). Research based on exploratory type questions (for example “what…”) will often be conducted using qualitative data (Ellis & Levy, 2009).

While the researcher is aware that the RDBMS environment and the role of the DBA are well-researched areas, the field of big data is continually evolving and being researched, and the introduction of the data scientist within the big data environment has to some extent muddied the waters (section 2.8). Murawski and Bick (2017) also noted that research to explore the skills required by the different roles within the big data environment is still exploratory to a large degree. The boundaries between the functions of the DBA, the functions of the data scientist, and the functions of other roles in a big data environment need further exploration. This exploration includes the technical skills required of the DBA to transition to a similar role within the big data environment.

3.4.2 Research assumptions

It is important to inform the reader of the assumptions made by the researcher in order to be open and transparent within any research (Ellis & Levy, 2009). Given the wide research area of big data in the last decade, the researcher assumes that the DBAs in this study will have had exposure to big data, or at least have a general understanding of the concept.

3.4.3 Research delimitations

Each research study has a scope, and it is important to bring to the reader's attention to the boundaries within which the study is conducted (Ellis & Levy, 2009). While this research is aiming to explore the difference in skills required between the DBA working in an RDBMS environment and a similar role within a big data environment, it is not the intention of the study to provide an in-depth technical assessment of the technologies identified.

3.4.4 Time horizons

The researcher must consider the time horizons for a study and hence deciding between longitudinal and cross-sectional studies (Gray, 2013:35). The intention of the study is not to explore what changes over time, given that technology within the big data environment is continuously evolving. It is also important to consider factors such as time-frame within which the research needs to be concluded as well as access to the resources taking part in the study (Gray, 2013:35). As such this study will be cross-sectional rather than longitudinal.

3.5 Data collection

A key component of research is the availability of data to the researcher, as this is one of the influencing factors of the research methodology and the type of research

conducted (Ellis & Levy, 2009). The interpretivist researcher uses data collection methods that will enable getting a thick description of the environment under study, driven by the respondent's own experience of reality (Tuli, 2010).

The primary data will be collected from the research participants in this study (Myers, 1997), namely DBAs. Interviews will be used as the method for data collection (Tuli, 2010).

3.5.1 Units of analysis

The UoA is an integral part of many studies, and often, but not always, the subject from whom this data is collected (Yin, 2006). In determining the UoA, it is important to understand what the study is focusing on and this can be at the organisational or individual level amongst others (Benbasat, Goldstein & Mead, 1987). The target population for this study will be all IT professionals with the job title of "database administrator or DBA" or a person who either has the primary task of administering databases but does not necessarily have the exact job title (section 2.3). Hence, the UoA within this study will be DBAs administering an RDBMS environment.

3.5.2 Sampling

Quantitative research commonly makes use of random sampling, also known as probability sampling (Marshall, 1996), and it is not uncommon for it to be used within qualitative research (Mays & Pope, 1995). Unlike the intention within quantitative studies to generalise the findings to the population, this study is not aiming at generalising the results. It would be time-consuming and inefficient to randomly sample individuals to see if they fulfil the role of a DBA. The DBA role is a specialist role within the IT environment (section 2.3), and given the sensitive nature of the data they manage in many organisations it might be difficult to gain access to a large enough sample to be considered a normal distribution.

Mays and Pope (1995) note that qualitative studies are costly and time-consuming. For these reasons this study will make use of the commonly used non-probability convenience sampling for selecting interviewees. One of the characteristics of non-probability sampling is identifying a group of people with specific skills or characteristics (Mays & Pope, 1995) and the DBA is a specialist role within the IT environment. In addition, individuals within this role fulfil a specific task and function and hence have specific knowledge and skills the researcher is aiming to access. There is a risk of gathering inferior data using convenience sampling (Marshall,

1996), but the researcher is working within the constraints of access to DBAs within organisations.

There is a fine balance between selecting the correct or ideal sample size of respondents for study. Too small sample sizes may result in not having enough data to explore the area under investigation properly or to make meaningful generalisations about the subject matter. Too big a sample size can hinder the researcher performing an extensive and in-depth analysis due to the sheer volume of data present (Kvale, 2011c:43). Qualitative research generally makes use of smaller sample sizes, unlike quantitative research which needs a sufficiently large sample size in order for it to be considered representative of the target population (Seale & Silverman, 1997). This study is qualitative and having a smaller sample size is considered acceptable. A sample size of approximately 15 interviews will be targeted (Kvale, 2011c:44).

3.5.3 Interviews

Qualitative studies can use a variety of data collection methods which include observation, interviews, surveys, documents, and texts (Kaplan & Maxwell, 2005). Interviews are also considered a valid data collection instrument in both quantitative and qualitative research methodologies (Seale & Silverman, 1997). Walsham (2006) states that interviews are an important tool in interpretive studies for understanding participants' views. It allows one to gain an understanding of the world from the respondent's perspective and their experiences (Kvale, 2011c:46). The interaction between the interviewer, the person asking the questions, and the interviewee, the person answering the questions, is in the form of a conversation with the desired outcome of knowledge being generated (Kvale, 2011b:7). For that reason, semi-structured interviews will be conducted. The research participants will be briefed as to the background of the study before-hand, and all interviews will be audio recorded with the consent of the interviewee. Recording the interview is acceptable given the richness of data generated through interviews within qualitative studies (Kaplan & Maxwell, 2005). "Compared to field notes of observational data, recordings and transcripts can offer a highly reliable record to which researchers can return as they develop new hypotheses" (Seale & Silverman, 1997:380).

3.5.3.1 Semi-structured interviews

It was decided to make use of "semi-structured interviews", due to it being a method suitable to gaining peoples' opinions on unexplored and or complex subject matters (Louise Barriball & While, 1994). Further probing questions were included for each

interview question to get as much information from the interviewee in answering the question (Turner, 2010). The semi-structured interview allows for the same set of open-ended interview questions to be asked of all respondents but with the flexibility to ask these questions in any order and probe additional avenues and allow for new avenues to emerge (Dearnley, 2005).

3.5.3.2 Interview consent (audio recordings)

Verbal consent will be sought to allow the researcher to record the interview. This consent will be captured on the researcher's smart-phone with the permission of the respondent. Recording the interviews allows the researcher not to be distracted by having to write down or remember what was said and more attention can be given to the actual answers provided and emotions displayed by the respondent (Kvale, 2011d:93).

3.5.3.3 Interview pseudonyms

Respondents will be given the guarantee that their identity will remain anonymous from any publications resulting from the study. Even though respondent identities will remain anonymous, it is still required that the researcher can identify recordings and transcriptions. The use of pseudonyms was considered and will be made use of, as it was done with great success by Dearnley (2005). Unlike Dearnley (2005) who allowed respondents to choose their own pseudonym, in this study the researcher himself will provide a pseudonym or number to each respondent to avoid any complications or delays in the interview process. Each respondent will be identified by a numeric value preceded by the letter "R", such as "R1". Only the researcher and the research supervisor at CPUT will be allowed to know the true identities of the respondents associated with the pseudonym.

3.5.3.4 Interview locations

Dearnley (2005) states that careful consideration is needed regarding the location of the interviews. It is imperative that the respondent is made as comfortable as possible about their surroundings and thus the researcher will allow respondents to choose a location in which they feel comfortable. The researcher's first preference will be to conduct the interview within the working environment of the research respondent.

3.5.3.5 Interview pilot

A pilot test of the interview questions will be conducted before the actual interview process starts to identify and correct any potential problem areas. One of the

reasons for such a pilot will be to identify if the respondent can answer the questions presented (Louise Barriball & While, 1994). It should be noted that the researcher will make an assessment as to the success of the interview pilot and will incorporate it into the data corpus if it is considered suitable.

3.5.3.6 Interview duration

The estimated duration of the interviews is approximately 30 minutes (DiCicco-Bloom & Crabtree, 2006). Dearnley (2005) reports that time becomes a factor to be seriously considered as the researcher requires the respondents to set aside time in the day on several occasions to participate in the study. Even though in this study the respondents will only be required to participate once in the interview, it is also something that will require careful consideration considering how busy DBAs can be.

3.6 Data analysis

3.6.1 Transcribing the interviews

All interviews will be transcribed verbatim by the researcher. Even though the interviews would not have any personal or sensitive information that could potentially be of concern to the respondents, it was decided the researcher would do the transcription of the audio recordings himself as this encourages early engagement with the data. The audio recordings will be securely safeguarded in a password protected zip file and will be discarded when it is acceptable to do so. Given the nature of the questions within the interview guide and the use of pseudonyms, there is no risk of any sensitive information emerging concerning the participant or their organisation.

3.6.2 Thematic analysis

The data will be analysed using the general inductive approach. The process adopted for the analysis of this data is known as “thematic analysis”. It “is a method for identifying, analysing, and reporting patterns (themes) within data” (Braun & Clarke, 2006:6). Thomas (2003) describes this approach as a systematic procedure that allows for research themes to emerge from the findings of complex raw data. Hence, no existing theories will be tested in this study, but theories will be allowed to emerge from the data analysed. A semantic approach to the thematic analysis will be used within this study. It is not the intention to identify the underlying factors that gave rise to the themes but merely to identify the patterns that have emerged from the data (Braun & Clarke, 2006).

3.6.2.1 Coding

Various methods for organising the keywords and key phrases into similar or the same bundle exist, whether by writing these codes onto pieces of paper and organizing them into heaps, referred to as “cutting and sorting”, or performing this action electronically (Ryan & Bernard, 2003). Care must be taken that the context of the code is not lost. This is done by including additional information relating to the respondent, as well as the location of the keyword or phrase in-text, and thus to the specific research sub-question to which it relates. There is no pre-defined coding scheme by which the researcher will code the data. Kaplan and Maxwell (2005) note that codes are generated in different ways which can include the researcher’s knowledge of existing literature as well as the codes arising from the text during data analysis. The researcher will perform the coding process electronically and will thus make use of MS Word to highlight the keywords and key phrases in the text. Thereafter, codes will be associated with the relevant keywords and key phrases identified. This will be an iterative process as the researcher becomes more familiar with the data.

3.6.2.2 Categorising codes and developing themes

This analysis will thus consist of both categorising the data and developing themes (Vaismoradi, Turunen & Bondas, 2013). Codes produced will be categorised into bundles that relate to the same concept according to those which arise from the interviews. The combining of codes is the process of working with or analysing the codes, rather than the interview transcript data itself (Braun & Clarke, 2006). Thereafter, the analysis will broaden to the categories, and similar categories will be grouped according to themes that emerge across these categories. The different phases making up the thematic analysis process will continuously be revisited and revised throughout the entire process in an iterative manner. No one process will ever be entirely complete before moving onto the next phase (Vaismoradi *et al.*, 2013).

The desired outcome of themes is to be able to identify specific patterns that emerge from the data under analysis and hence relate this back or provide an answer to the research question(s) to be answered (Braun & Clarke, 2006). Due to the subjective nature of qualitative research, the number of occurrences of a theme within the data is not an indicator of its importance (Braun & Clarke, 2006). This is left to the judgement of the researcher as to how best such a theme relates to answering the research question(s).

3.7 Ethics and consent

According to Orb *et al.* (2001:93) “Ethics pertains to doing good and avoiding harm”. Kaplan and Maxwell (2005) thus note that a researcher must be cognizant of the fact that interviews can be considered an intrusive process. While interviews allow for free conversations between the researcher and the participant, researchers’ must be aware of the purpose of the research (Orb *et al.*, 2001) and the researcher must thus adhere to the study’s interview guide.

Even though the researcher will firstly get the needed approvals from management within the organisation, it is still required that the respondents’ participation in the study be noted as voluntary. It is also important that the manner in which the researcher is introduced to the respondents in the study be conducted ethically and correctly (Orb *et al.*, 2001). Thus, the researcher should be aware of this even though it can be potentially difficult getting access to DBAs.

Joe *et al.* (2016:6) note the “golden rule” within research is for participants to consent to the research. This consent entails providing them with enough information relating to the research, what the information will be used for, that their anonymity is guaranteed and to what extent their information will be confidential.

All respondents will thus take part on a voluntary basis and will be free to withdraw from the study at any stage. Respondents will be provided with the background to the research before starting with the interview. Their confirmation that they understand what the research is about, and their voluntary participation will be documented in an interview consent form (see APPENDIX B). In addition, prior to the researcher asking the first interview question, an audio recording stating that they voluntary consent will be made in their presence.

It is of utmost importance for the researcher to ensure that participants within the study remain anonymous. DiCicco-Bloom and Crabtree (2006) state that even though by revealing information from the interview process may improve the working environment of the respondents, it is possible that the participant is subject to conflict within their working environment because of this. Hence anonymity is extremely important to prevent harm in any circumstances. To this end, the respondent will be given the opportunity to state their desire to either be named in research publications or remain anonymous. This is aligned with the researcher providing every respondent with a pseudonym, and confirmation of the respondent’s acceptance to be interviewed using the given pseudonym rather than by name. Further to this, the

respondents will also be requested not to supply any names of organisations at which they are or have been employed as well as not naming of individuals. The researcher will further anonymise any such information in the transcripts should this request accidentally not be followed.

Lastly Orb *et al.* (2001) state that the responsibility ultimately lies with the researcher to ensure that the participants are protected irrespective of various clearances provided to proceed with data collection. In this regard, the researcher will be cognisant and adhere to the four ethical issues raised by DiCicco-Bloom and Crabtree (2006:319) while conducting the actual interview process:

(i) "reducing the risk of unanticipated harm;"
(ii) "protecting the interviewee's information;"
(iii) "effectively informing interviewees about the nature of the study, and"
(iv) "reducing the risk of exploitation."

3.8 Delineation

This study focuses on the technical rather than the business or social re-skilling requirements of DBAs. Face-to-face interviews will be conducted only with research participants where geographically possible. The research will be conducted with DBAs working within multi-national organisations within South Africa. Re-skilling in other areas affected by the requirements of a big data environment will not be considered and can be investigated in future research.

3.9 Summary of Chapter Three

Chapter Three provided an insight as to how the researcher intends to proceed with the research and the foundations upon which it is based. This was presented in relation to the underlying philosophical paradigms, research methodologies, research approaches and research design followed and how they relate to each other.

With any research, it is important to consider the primary philosophical paradigms namely subjectivist or objectivist. This is because everyone, including the researcher, experience the world and events that occur therein in a different way. These experiences are also influenced by the researcher's own interest in the subject matter under study. As such, it is important to identify the underlying core assumptions relating to ontology (reality) and epistemology (knowledge) adopted by the researcher as this also influences the research methodology, research design and data collection methods to be incorporated within the study (section 3.2). The

researcher considers that the nature and knowledge of reality for the respondents in this study is not devoid of their own experiences within their working environment. As such, this research follows the constructionist ontological and interpretivist epistemological stances which are closely aligned.

The philosophical stances adopted in turn have influenced the research approach in that a qualitative research methodology has been adopted. The environment of each DBA is different and is thus difficult to understand through the scientific enquiry of quantitative studies. Furthermore, because it is relatively unknown as to what skills the DBA would require enabling them to transition to a big data environment, an inductive approach will be adopted. For that reason, an exploratory research design was adopted for this study using interviews as a data collection method to obtain primary data. The literature will also be examined to substantiate the background of the study and to support the discussion of its findings. These interviews are semi-structured interviews using open-ended questions to allow the researcher an understanding of how such DBAs are experiencing and considering a transition to a big data environment.

Data analysis will be done using thematic analysis. Semantic level coding will be done electronically using MS Word, and the relevant codes will be categorised and themed to answer the research questions within this study.

Chapter Four will present the actual field work relating to the data collection and data analysis processes as outlined in this chapter.

CHAPTER FOUR: DATA COLLECTION AND DATA ANALYSIS

4.1 Introduction

Chapter Three described the research design and methodology implemented in this study. This chapter provides a detailed report of the actual process undertaken, including an initial analysis of the data that emerged from the data collection process.

4.2 Data collection

The data collection was conducted using semi-structured interviews. The interviews were conducted face-to-face with all respondents. It took approximately six months to conduct the fifteen interviews, with the first interview on the 14th September 2017 and the last interview being conducted on 28th March 2018. A detailed description of the entire process is described in the following sections.

4.2.1 Participant sampling

This research made use of probability sampling, namely convenience sampling (section 3.5.2). While a sincere attempt was made to gain access to DBAs through their company structures, the researcher was also required to make use of snowball sampling in some instances. Snowball sampling is a convenience sampling method which relies on referrals from one respondent to another, and it is used for research populations that are difficult to access (Johnston & Sabin, 2010). Five respondents in this study were found using the snowball sampling method.

All respondents selected for the study either explicitly had a variation of the title “database administrator” at their respective companies or had a role in administering a database. Initially, 16 respondents were scheduled to take part in the study, from six different companies within various sectors in South Africa. Eventually, only 15 interviews were conducted.

Respondents were located in the two major centres of South Africa, Cape Town and Johannesburg and the researcher travelled to Johannesburg to conduct the interviews in person with the respondents located there.

4.2.2 Interview locations

The interviews were conducted at various locations, with the researcher giving the respondent the choice of choosing their preferred location where they might feel comfortable. Nine interviews took place at the organisations’ offices at which the

respondents were employed, while the other six respondents felt more comfortable to meet the researcher at either a restaurant or coffee shop which was conveniently located for them.

4.2.3 Interview process

The interview process is now described in detail.

4.2.3.1 Interview introduction

The interview with each respondent started with the researcher providing an overview of the research being conducted by describing the problem statement to be addressed as well as the aim of the research. Any questions following on from this introduction were addressed first before starting with the interview. Kvale (2011a:55) states that it is of utmost importance that the respondent be made as comfortable as possible and be provided with a detailed introduction to the research being undertaken, as this initial stage of the interview will determine the success of the rest of the interview.

Once the introduction was complete, the researcher assigned each respondent a pseudonym by which they were identified for the duration of the interview. These pseudonyms were R1 to R15.

4.2.3.2 Interview consent

Each respondent was requested to complete a consent form stating their voluntary participation in the research. This also gave them the opportunity to indicate their willingness to be named or remain anonymous in any publications resulting from this study (see APPENDIX B for a signed interview consent form).

4.2.3.3 Interview audio recordings

The interview process was started by asking the respondent if they grant permission for the interview to be recorded. The recording of the interview took place on the researcher's smartphone, and this process of asking permission was also recorded. The smartphone was placed on the table between the respondent and the researcher so that the respondent could see everything that was being done and that the process was open and transparent. If the respondent declined this request, then the recording device was switched off in view of the respondent. One respondent (R8) declined the invitation to have the interview recorded. The respondent's reasoning for this was for fear of saying anything that could negatively reflect on their employer. Even though the researcher assured the respondent that none of the

questions can in any way be considered of such a nature that it could jeopardise the respondent's employment or reflect negatively on the employer. The respondent still refused, and a manual recording was done. Short-hand notes were made by the researcher on the interview guide. This led to a significantly longer interview than expected and resulted in the last research sub-question being rushed.

Even though the rest of the respondents allowed the interview to be recorded, four respondents needed reassuring that they will remain anonymous and that nothing within the interview will negatively reflect upon their employer. Upon hearing that the executive management of the company had given approval for the interview to take place, they felt more assured. Only one respondent specifically asked that the interview not be shared with the employer.

A copy of the interview recordings was stored in a password protected zip file and uploaded to Google Drive for safe-keeping. Given that there were no company names or IP, or personal information of the respondents included on the audio files, the researcher found the measures implemented for safe-keeping as sufficient.

4.2.3.4 Interview questions

All respondents were asked the same set of interview questions. Additional probing questions were developed for each of the interview questions to provide more insight into the interview question being asked. This probing was found to be especially useful and provided a rich data corpus for the research (see APPENDIX C for the full interview guide).

4.2.3.5 Interview debriefing

At the end of the interview, the researcher provided the respondent with the opportunity to ask any questions or voice any concerns they might have, as well as the experience of the interview process just completed. Any feedback about the interview and questions was recorded accordingly. This is known as debriefing which occurs after the interview process (Kvale, 2011a:56) as the interview process can be an anxious experience for some individuals.

One respondent (R6) used the opportunity after the interview to get the researcher's view on the subject matter being researched. This was a casual conversion and did not form part of the coding process.

4.3 Data analysis

4.3.1 Transcribing the interviews

Verbatim transcriptions of the audio recordings were made. The audio recordings were then listened to twice: firstly, to verify that the transcription is a correct reflection of the audio; and secondly, to try and capture the context in which the respondent answered the question. The audio recordings were transcribed into an MS Word document generating a total of 136 pages of transcriptions for the 15 interviews. Transcribing the audio recordings was a time-consuming process given the different environments in which the interviews were conducted. Background noise in some of the interview locations made it difficult to hear clearly what was being said by some of the respondents. The format of the interview transcripts is shown in Figure 4-1.

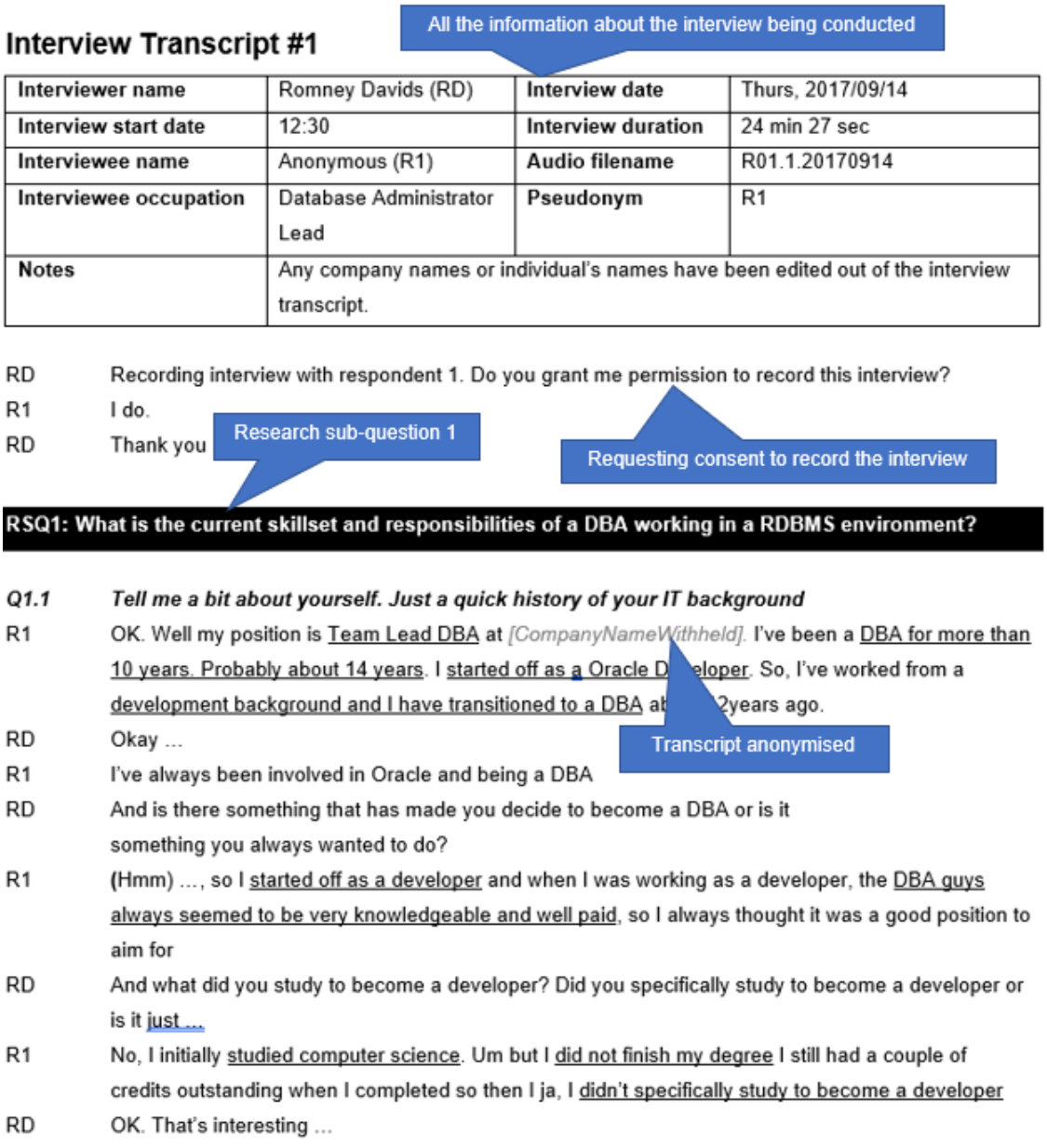


Figure 4-1: Snapshot of transcript for respondent R1

General information that would allow the researcher to identify the interview was captured in the header of each interview transcript, e.g. respondents were indicated by the pseudonym (R1-R15) in the transcript, and the interviewer was denoted as “RD”.

4.3.2 Getting familiar with the data

The entire data corpus consisted of 15 interviews, each covering interview questions relating to the three research sub-questions. Research sub-question one consisted of five interview questions, research sub-question two consisted of four interview questions and research sub-question three consisted of two interview questions.

The approach taken to identify keywords and key phrases was to cycle through the same interview question for every transcript before continuing to the next interview question. For example, interview question 1.1 for all the respondents was examined before continuing to the next interview question (i.e. interview question 1.2) for each respondent.

The original interview transcript (see Figure 4-1) was copied to a new document which contained two columns as shown below in Figure 4-2. The first column named “Transcript” contains the transcribed interview and the second column “Notes” was used for note-making as part of the first phase of getting to know the data. These were not the final codes but merely short summaries and ideas of what the keywords and key phrases meant when initially getting to know the data. This was done to help the researcher once the actual process of creating codes started. Some of these notes did go on to become actual codes. This process was followed for each interview question for each respondent until all the transcripts were complete. For each question, various keywords and key phrases were identified which could aid in answering the research sub-question. These keywords and key phrases were highlighted in MS Word as shown in Figure 4-2.

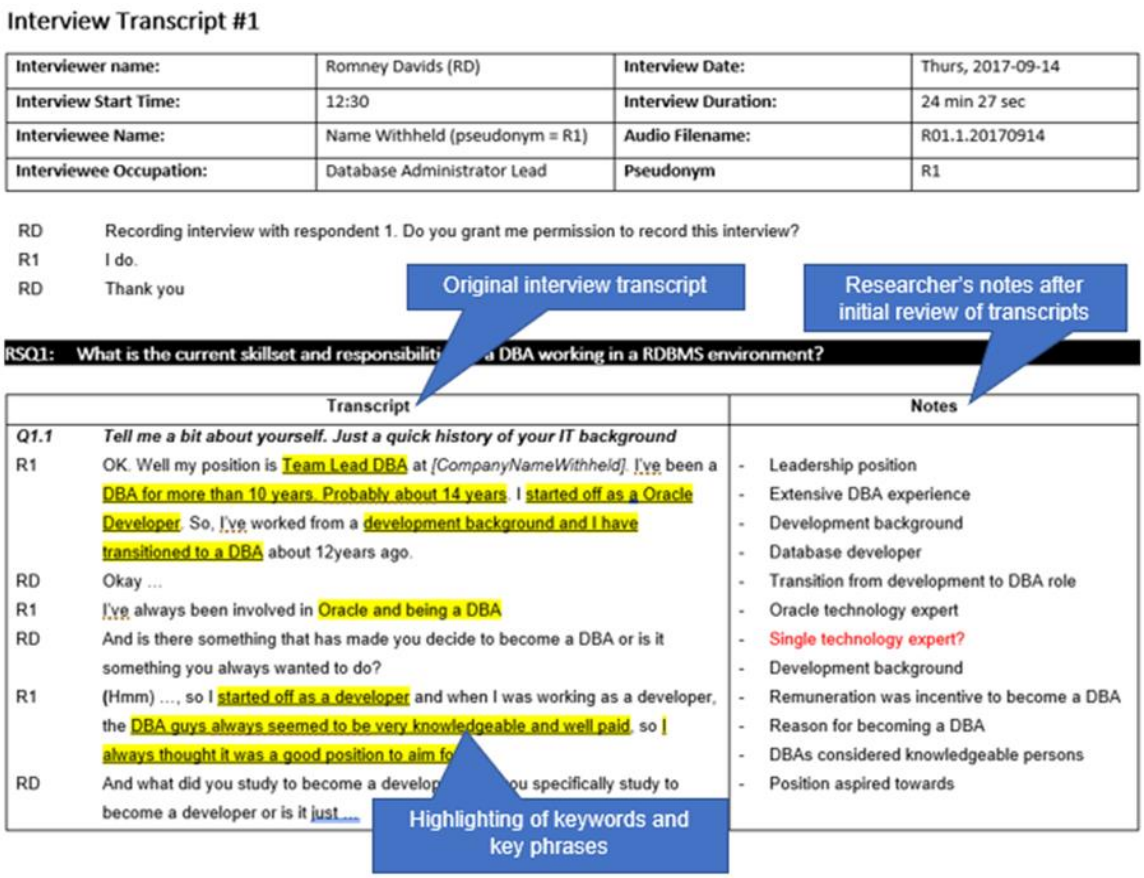


Figure 4-2: Snapshot of getting familiar with transcribed interview

4.3.3 Generating the initial codes

There was no pre-defined list of codes used by the researcher to code the data, but an open coding method was used for creating new codes as the interview text was being examined (Maguire & Delahunt, 2017).

As part of the process of finding codes, the interview transcripts for all respondents by research sub-question was input into the Word Cloud application "TagCrowd" (https://tagcrowd.com/) to identify the top 50 words used by all respondents to aid formulation of codes. Input parameters for generating the word clouds were as follows - similar words were grouped together, the top 50 words were shown, and some common words were included in the ignore list across all research sub-questions. These included words such as "hmm", "think", "going", "ja", "dba", "okay" etc.

Codes were created within the MS Word document using the "comments" feature using the following pattern. Each item was delimited by a colon ":".

Format: *<respondent-pseudonym>:<question-number>:<code>*

Example: *R1:1.1:Position aspired towards*

An example of this would be for respondent "**R1**", interview question number "**1.1**" and code being "**Position aspired towards**" as shown in Figure 4-3:

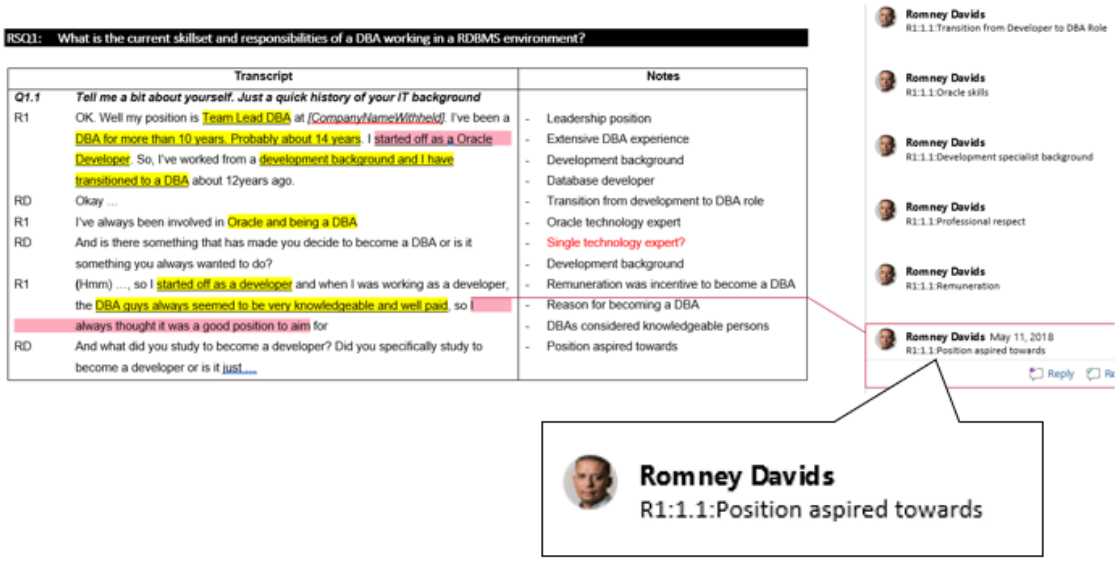


Figure 4-3: Generating codes from keywords and key phrases for respondent 1

The codes, along with verbatim text, were extracted from the MS Word document using a Visual Basic script embedded as a macro. The extracted information was then transferred to a comma-separated values (CSV) file and imported into an MS Excel workbook. The initial coding process produced a total of 1375 keywords.

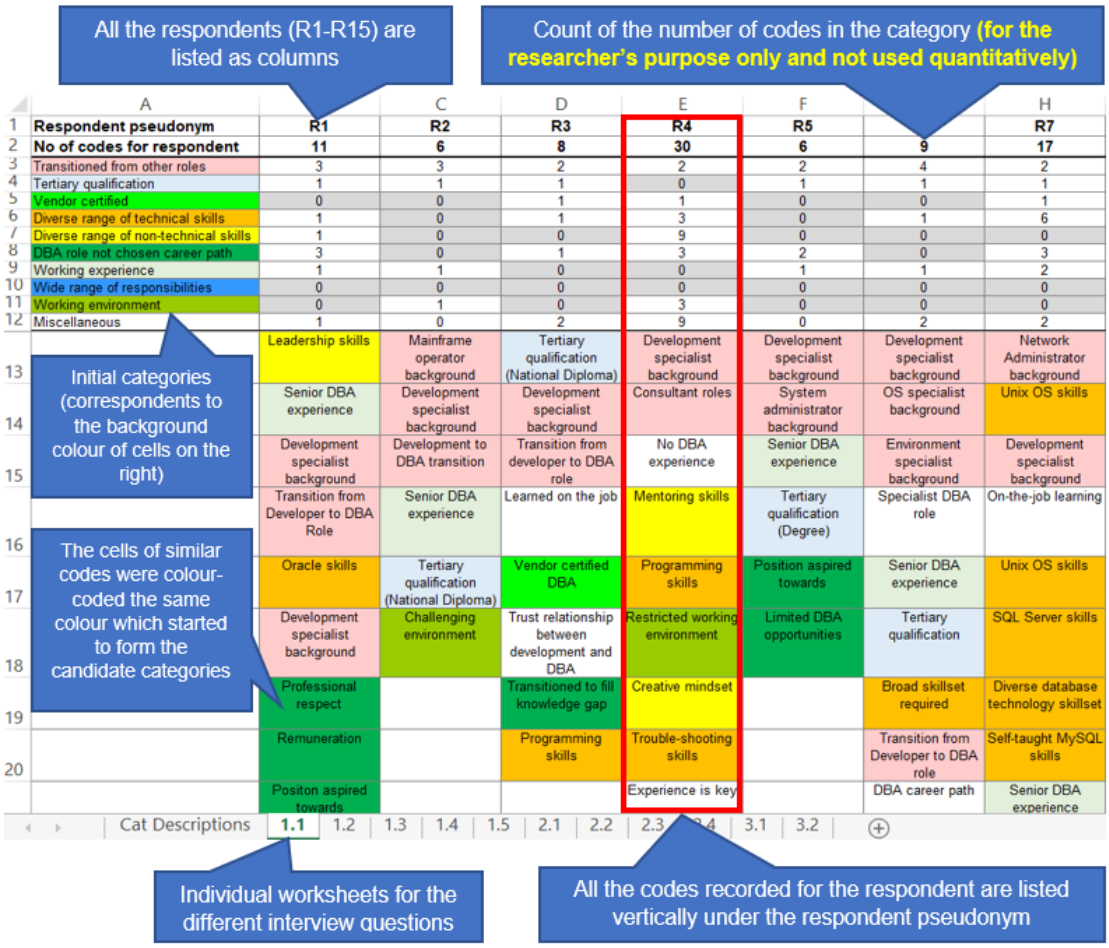
	A	B	C	D	E	F
1	Resp	Que	Verbatim Text	Code	Coder	Date Coded
2	R1	1.1	Team Lead DBA	Leadership Skills	Romney Davids	2018-05-08 23:33
3	R1	1.1	DBA for more than 10 years. Probably about 14 years	Senior DBA experience	Romney Davids	2018-05-08 23:34
4	R1	1.1	started off as a Oracle Developer	Development specialist background	Romney Davids	2018-05-08 23:34
5	R1	1.1	development background and I have transitioned to a DBA	Transition from Developer to DBA Role	Romney Davids	2018-05-08 23:35
6	R1	1.1	Oracle and being a DBA	Oracle skills	Romney Davids	2018-05-11 20:52
7	R1	1.1	so I started off as a developer	Development specialist background	Romney Davids	2018-05-11 20:53
8	R1	1.1	DBA guys always seemed to be very knowledgeable	Professional respect	Romney Davids	2018-05-11 20:53
9	R1	1.1	DBA guys always seemed to be very knowledgeable and well paid	Remuneration	Romney Davids	2018-05-14 19:48
10	R1	1.1	I always thought it was a good position to aim	Position aspired towards	Romney Davids	2018-05-11 20:56
11	R1	1.1	I did not finish my degree	Incomplete tertiary qualification/degree	Romney Davids	2018-05-11 20:57
12	R1	1.1	I didn't specifically study to become a developer	Studies not aimed at becoming developer	Romney Davids	2018-05-11 20:58
13	R1	1.2	in the mornings I go through my emails to see if there are any messages from our alert system. We have a system that checks alert logs and makes sure that it does health checks on all of our databases	Automated monitoring jobs	Romney Davids	2018-05-11 20:59
14	R1	1.2	in the mornings I go through my emails to see if there are any messages from our alert system. We have a system that checks alert logs and makes sure that it does health checks on all of our databases	Set morning routine	Romney Davids	2018-06-03 22:55

Figure 4-4: Codes and verbatim text extracts in an MS worksheet

4.3.4 Searching for categories

After all the codes were extracted from the MS Excel workbook (see Figure 4-4), a new workbook was created with individual worksheets for every interview question. Respondents were listed in columns, and all the codes extracted for the respondent for each interview question were listed under the respondent pseudonym.

Figure 4-5 shows the cells for codes relating to the same or a similar concept were colour-coded, and the various categories started to emerge. Once the researcher was satisfied that a category was saturated, the category was then given an initial name and description. The cell for the description was given the same colour. A Visual Basic script was embedded in the worksheet to supply a count of the number of codes that belonged to a category. This was done for the researchers own informational purposes and not to conduct the data analysis quantitatively. This process was followed for all the interview questions across the research sub-questions.

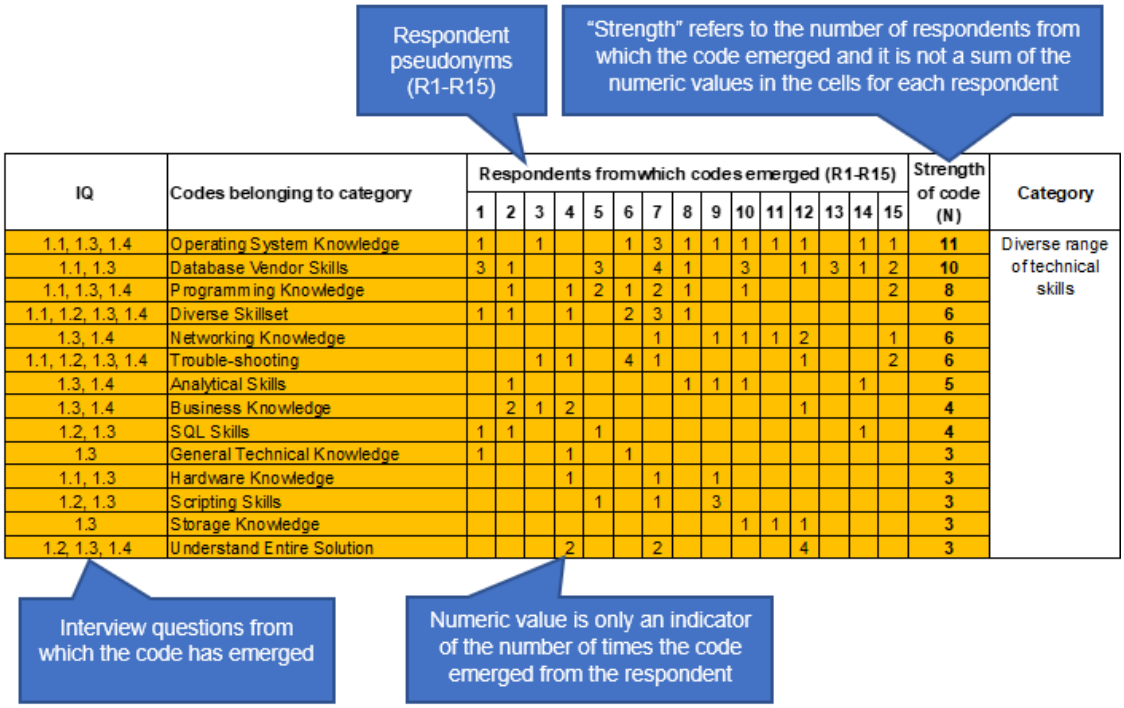


	A		C	D	E	F		H
1	**Respondent pseudonym**	**R1**	**R2**	**R3**	**R4**	**R5**		**R7**
2	**No of codes for respondent**	**11**	**6**	**8**	**30**	**6**	**9**	**17**
3	Transitioned from other roles	3	3	2	2	2	4	2
4	Tertiary qualification	1	1	1	0	1	1	1
5	Vendor certified	0	0	1	1	0	0	1
6	Diverse range of technical skills	1	0	1	3	0	1	6
7	Diverse range of non-technical skills	1	0	0	9	0	0	0
8	DBA role not chosen career path	3	0	1	3	2	0	3
9	Working experience	1	1	0	0	1	1	2
10	Wide range of responsibilities	0	0	0	0	0	0	0
11	Working environment	0	1	0	3	0	0	0
12	Miscellaneous	1	0	2	9	0	2	2
13		Leadership skills	Mainframe operator background	Tertiary qualification (National Diploma)	Development specialist background	Development specialist background	Development specialist background	Network Administrator background
14		Senior DBA experience	Development specialist background	Development specialist background	Consultant roles	System administrator background	OS specialist background	Unix OS skills
15		Development specialist background	Development to DBA transition	Transition from developer to DBA role	No DBA experience	Senior DBA experience	Environment specialist background	Development specialist background
16		Transition from Developer to DBA Role	Senior DBA experience	Learned on the job	Mentoring skills	Tertiary qualification (Degree)	Specialist DBA role	On-the-job learning
17		Oracle skills	Tertiary qualification (National Diploma)	Vendor certified DBA	Programming skills	Position aspired towards	Senior DBA experience	Unix OS skills
18		Development specialist background	Challenging environment	Trust relationship between development and DBA	Restricted working environment	Limited DBA opportunities	Tertiary qualification	SQL Server skills
19		Professional respect		Transitioned to fill knowledge gap	Creative mindset		Broad skillset required	Diverse database technology skillset
20		Remuneration		Programming skills	Trouble-shooting skills		Transition from Developer to DBA role	Self-taught MySQL skills
		Positon aspired towards			Experience is key		DBA career path	Senior DBA experience

Figure 4-5: Colour coding initial codes into categories

After the initial colour-coding of the cells to map similar codes into categories, all codes with the same colour cells were copied into a new MS Excel workbook with all those codes listed below one another (see Figure 4-6). Given the extensive list of initial codes that emerged, they were further summarised or collapsed during this phase by combining codes of the same meaning for the various respondents. Codes that did not match any of the criteria needed to answer the research sub-questions were placed in a "**Miscellaneous**" category to avoid misreporting the data.

The same codes that emerged across different interview questions were combined, and the IQ (interview question) column was updated with the interview question number (see Figure 4-6). For example, the code "**Operating System Knowledge**" emerged from 11 respondents (R1, R3, R6-R12, R14, R15) across interview questions 1.1, 1.3 and 1.4. This was repeated until a final list of codes associated with each category was produced.



IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1.1, 1.3, 1.4	Operating System Knowledge	1		1			1	3	1	1	1	1	1		1	1	**11**	Diverse range of technical skills
1.1, 1.3	Database Vendor Skills	3	1			3		4	1		3		1	3	1	2	**10**	
1.1, 1.3, 1.4	Programming Knowledge		1		1	2	1	2	1		1					2	**8**	
1.1, 1.2, 1.3, 1.4	Diverse Skillset	1	1		1		2	3	1								**6**	
1.3, 1.4	Networking Knowledge							1		1	1	1	2			1	**6**	
1.1, 1.2, 1.3, 1.4	Trouble-shooting			1	1		4	1					1			2	**6**	
1.3, 1.4	Analytical Skills		1						1	1	1				1		**5**	
1.3, 1.4	Business Knowledge		2	1	2								1				**4**	
1.2, 1.3	SQL Skills	1	1			1									1		**4**	
1.3	General Technical Knowledge	1			1		1										**3**	
1.1, 1.3	Hardware Knowledge				1			1		1							**3**	
1.2, 1.3	Scripting Skills					1		1		3							**3**	
1.3	Storage Knowledge										1	1	1				**3**	
1.2, 1.3, 1.4	Understand Entire Solution				2			2					4				**3**	

Figure 4-6: Codes collapsed across interview questions

Figure 4-6 is only an illustration of how the codes for the respondents were organised into categories. An expanded list of these codes can be found in APPENDIX G.

4.3.5 Searching for themes

This phase of the analysis process started with the categories identified after collating codes with similar meanings together (see Figure 4-5). Twenty-seven categories were identified in the previous phase. Some categories were quite extensive relative to the number of codes they contained given the granular nature of the coding.

The understanding of the various categories arose through the researcher interpreting and writing up of the findings for each category. Even though the initial coding and categorisation of codes were complete, the mapping of codes to categories was still an iterative process to ensure that codes fit together coherently

within each category. In some instances, codes were moved to other categories after reviewing the verbatim text related to that code to ensure the context of the code remained intact. This iterative process of reviewing the various categories helped the researcher to start grouping categories together for the emergence of seven themes.

4.3.6 Final list of categories and themes

The final list of categories and themes that emerged is shown below in Table 4-1 (Column "**IQ**" refers to the interview questions from which the codes for the categories were derived).

Table 4-1: Overview of categories and themes

IQ	Categories	Themes
1.1, 1.2, 1.3, 1.4	Diverse range of technical skills	Skills and responsibilities (RDBMS environment)
	Diverse range of non-technical skills	
	Wide range of responsibilities	
	Automation of tasks	
	DBAs have non-dba functions	
	Interface with a variety of stakeholders	
	Working environment	
1.1	Transition from other roles	Background and education
	Tertiary qualifications	
	DBA role not chosen career path	
1.1, 1.3, 1.4, 1.5, 2.1, 2.3, 2.4, 3.1	Attitude towards tertiary qualifications	Training and upskilling
	Attitude towards vocational training	
	Importance of on-the-job experience	
	Methods of upskilling	
2.1, 2.2, 2.3, 2.4, 3.1	Shortage of big data exposure	Adoption of big data
	Limited big data knowledge	
	Attitude towards big data	
	Attitude towards RDBMS	
2.1, 2.2, 2.3, 2.4	Expansion of technical skills	Skills and responsibilities (Big data environment)
	Less technical skills	
	DBA role and responsibilities in big data	
2.1, 2.4, 3.1, 3.2	Skills development opportunities	Big data career pathing
	Limited big data opportunities	
	Future career aspirations	
2.3, 3.1	Ability to transition	Transitioning to a big data role
	Re-skilling	
	Attitude towards being certified in big data	

4.4 Summary of Chapter Four

Chapter Four provided a detailed overview of the data collection process that was applied in this study. Fifteen respondents from six organisations were interviewed, selected through using non-probability sampling methods, convenience and snowball sampling. Semi-structured interviews were conducted face-to-face with all respondents over a period of six months, which required the researcher to travel from Cape Town to Johannesburg for some of the respondents.

Included in this chapter was the data analysis process applied in respect to thematic analysis and was described in detail in this chapter. The analysis process included transcribing the recorded interviews, generating the initial codes from highlighted keywords and key phrases, the development of 27 categories and resulting in the emergence of 7 themes. The findings for the themes and underlying categories are presented in Chapter Five.

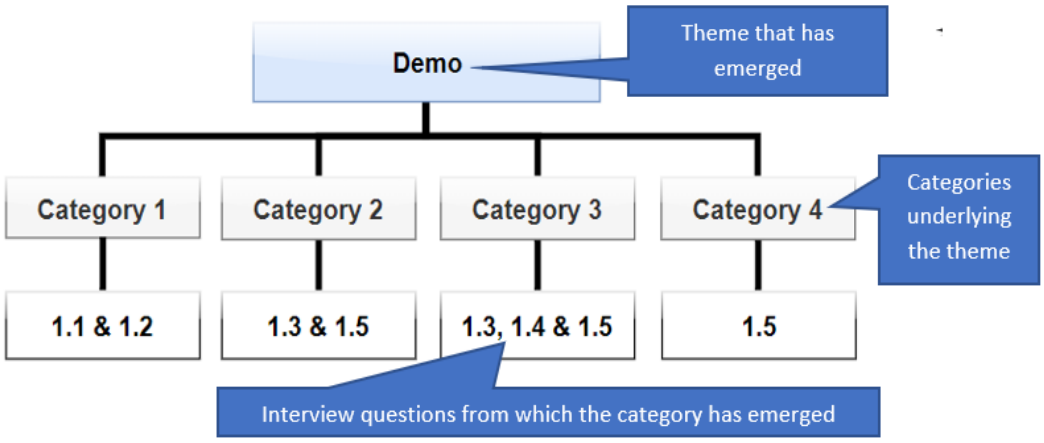
CHAPTER FIVE: RESEARCH FINDINGS

5.1 Introduction

Chapter Four provided an in-depth account of the actual data collection and data analysis processes followed in this study. It also presented the categories and subsequent themes that emerged from the data collected during the interview process. This study aims to explore the difference in skill set between a DBA administering an RDBMS and similar position in a big data environment.

5.1.1 Overview of the presentation of the findings for each theme

Each theme will be introduced using a diagram as shown below. The figure shows the theme being reported, the underlying categories from which the theme has emerged as well as the interview questions from which the codes for the categories were derived.



The researcher would like to present the following for the reader's attention:

(i) The data corpus consists of fifteen completed interviews.

(ii) At the end of each category, a table will be presented to show the distribution of responses for each code amongst the respondents. (See the example table below).

(iii) The numeric values in the table under the respondent pseudonyms (R1-R15) were only used to show the number of times the code appeared for the respondent. It was not the intention to interpret the data quantitatively.

(iv) Given the diverse number of codes within many categories, only codes that have three or more respondents will be reported. This will be shown as the "**strength**" of the code.

(v) References to the verbatim text quotes as reported by respondents will be included in the findings. Short passages may be included in-text with the

reporting of the findings while longer passages will be presented after the paragraph. Below is an example of how verbatim text quotes will be presented for respondent R1 from page 100 in the interview transcript document:

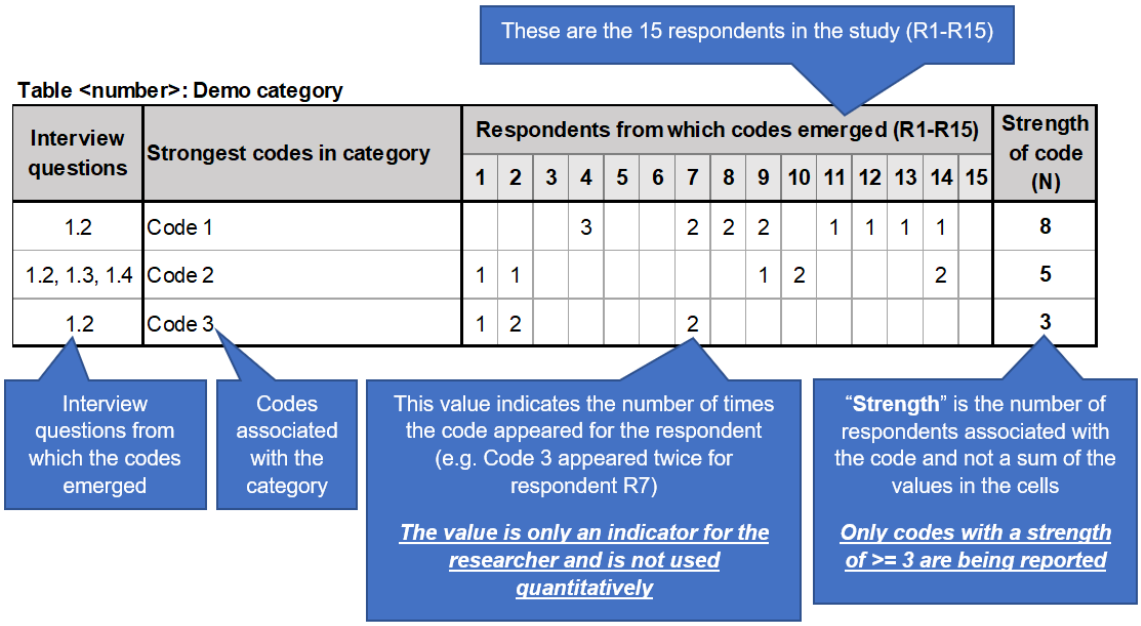
(R1) *"verbatim text quoted by respondent"* ***(100)***

(vi) The interview transcripts of all 15 respondents were combined into a single document hence high page numbers.

(vii) In order to avoid misrepresenting the data, the researcher would like to bring to the reader's attention the following categories which were not included in the findings of this study:

- Views on open-source software (*category does not aid in answering any of the research questions)*
- Views on proprietary software *(none of the codes in the category meets the minimum strength of three as noted in point iv above)*
- Understanding of big data *(none of the codes in the category meets the minimum strength of three as noted in point iv above)*

Example illustration of the table to be presented at the end of each category:



5.1.2 Themes and categories

Seven themes emerged from the thematic analysis process. Each of the themes presented below consisted of several categories which resulted from the concatenation of several codes. The categories that formed the themes will be presented in detail under each corresponding theme heading.

The themes are as follows:

(i) Skills and responsibilities (RDBMS environment).
(ii) Background and education.
(iii) Training and upskilling.
(iv) Adoption of big data.
(v) Skills and responsibilities (Big data environment).
(vi) Big data career pathing.
(vii) Transitioning to a big data role.

5.2 Theme - Skills and responsibilities (RDBMS environment)

The *"Skills and responsibilities (RDBMS environment)"* theme describes the current role of the DBA working in an RDBMS environment. The theme presents the diverse range of skills that DBAs reported, either because of their prior working experience in other roles or because of it being a requirement of the DBA role. Both technical and non-technical skills are reported. In addition to skills, this theme also reports on the responsibilities and functions within the DBA role, within their own department and across the organisation. Figure 5-1 below provides an overview of how this theme emerged.

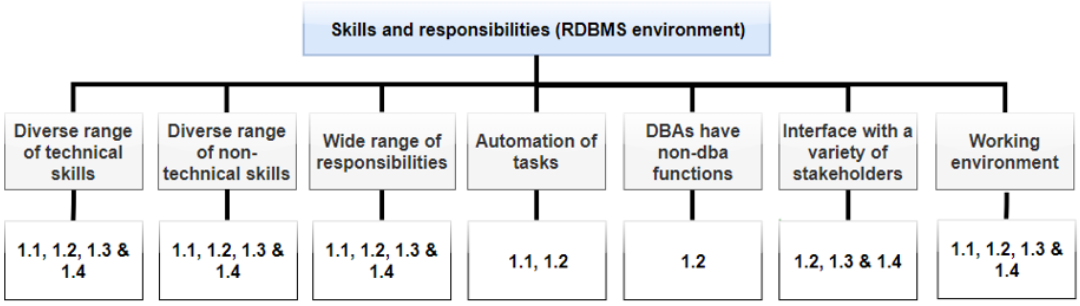


Figure 5-1: Skills and Responsibilities (RDBMS environment) theme

5.2.1 Diverse range of technical skills category

The range of technical skills reported by respondents to enable them to perform their daily duties as a DBA was quite diverse. The consensus amongst respondents was that the technical skills required by a DBA should not only focus on the management of the database software but should also include skills required to manage or administer the entire database solution. Respondents reported that the reason for such a diverse range of skills was because the DBA's role did not only cover the management of databases, but they also needed to have at least a basic understanding or have knowledge of the underlying infrastructure components which support the database. It also included surrounding systems that the database supports.

Findings showed strong support for having knowledge of, or having a basic understanding of, the various aspects relating to the supporting infrastructure underlying the database solution such as operating systems, networking, storage and hardware. Strong support was shown by eleven respondents who reported knowledge of the operating system (R1, R3, R6-R12, R14, R15), six respondents reported having at least a basic knowledge of networking (R7, R9-R12, R15), three respondents reported knowledge of hardware (R4, R7, R9) and three respondents reported knowledge relating to storage (R10-R12). These skills relating to the underlying infrastructure were further highlighted by three respondents explicitly stating that DBAs have a requirement to understand the entire solution (R4, R7, R12). The level of knowledge and ability required varied amongst respondents. Some respondents reported having such knowledge as advantageous while some respondents found it to be a requirement when wanting to understand the performance of their database. Respondents R1 and R7 spoke about the underlying components:

> ***R1:*** *"…you need some kind of OS skills, for us most of our work is done in Linux or Unix so those, the OS skills are important…"* ***(5)***
> ***R7: "…****When it comes to the networking side, I think you know sort of on a small sort of you know I know what IP addressing is…"* ***(26)***

The core focus of a DBA is to ensure that they are proficient in all the skills needed to manage and administer the database systems they must support. These skills also include the ability to troubleshoot problems that occur regularly and being proficient in SQL. Ten respondents reported the need for skills to manage the database vendor software (R1, R2, R5, R7, R8, R10, R12-R15), six respondents reported troubleshooting skills (R3, R4, R6, R7, R12, R15) and four respondents reported an ability in SQL (R1, R2, R5, R14). Respondents R5 and R6 spoke about knowledge of the database software and troubleshooting skills:

> ***R5:*** *"…(Hmm) … Ja I think knowledge of the type of the database that they using … (hmm) … they have to have pretty good knowledge of that…"* ***(63)***
> ***R6:*** *"… making sure that it is running efficiently, if there is any bottle-necks that you sort of get to it that you can find out what it is and what is causing it (hmm) … most cases its users running exceptionally bad SQL that's causing the problem …"* ***(76)***

The skill set of respondents did not only focus on database management and its underlying components, as DBAs must also have knowledge in areas not directly related to the management of the database software. Eight respondents showed strong support for having programming knowledge (R2, R4-R8, R10, R15) while three respondents also included scripting skills (R5, R7, R9). In addition, five respondents reported the need for analytical skills (R2, R8-R10, R14) while four respondents included the need to have business knowledge (R2-R4, R12). These skills were not a requirement to administer the database software as such but were needed in the wider role performed by the DBA. These skills are needed when DBAs must perform troubleshooting or assist different stakeholders within the organisation such as developers and analysts. Respondent R15 spoke about programming skills:

> ***R15:*** *"…It becomes a bonus if you have let's say programming skills because sometimes the developers or the client will say that the database is slow and maybe they are getting … an error …"* ***(205)***

Six respondents (R1, R2, R4, R6-R8) reported that DBAs need a diverse skill set while three respondents (R1, R4, R6) also reported the need to have general technical knowledge.

Table 5-1: Diverse range of technical skills category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1, 1.3, 1.4	Operating System Knowledge	1		1			1	3	1	1	1	1	1		1	1	**11**
1.1, 1.3	Database Vendor Skills	3	1			3		4	1		3		1	3	1	2	**10**
1.1, 1.3, 1.4	Programming Knowledge		1		1	2	1	2	1		1					2	**8**
1.1, 1.2, 1.3, 1.4	Diverse Skillset	1	1		1		2	3	1								**6**
1.3, 1.4	Networking Knowledge							1		1	1	1	2			1	**6**
1.1, 1.2, 1.3, 1.4	Trouble-shooting			1	1		4	1					1			2	**6**
1.3, 1.4	Analytical Skills		1						1	1	1				1		**5**
1.3, 1.4	Business Knowledge		2	1	2								1				**4**
1.2, 1.3	SQL Skills	1	1			1									1		**4**
1.3	General Technical Knowledge	1			1		1										**3**
1.1, 1.3	Hardware Knowledge				1			1		1							**3**
1.2, 1.3	Scripting Skills					1		1		3							**3**
1.3	Storage Knowledge										1	1	1				**3**
1.2, 1.3, 1.4	Understand Entire Solution				2			2					4				**3**

5.2.2 Diverse range of non-technical skills category

Non-technical or soft skills play an essential part in the range of skills required by a DBA. All the respondents felt that these skills were no less important than the wide variety of technical skills required to perform as a DBA. Again, the range of non-technical skills was extensive and ranged from the ability to be able to communicate effectively with various stakeholders within and outside of the organisation to skills such as self-motivation, teamwork, time-management skills including a range of other skills.

The requirement for DBAs to communicate effectively was evident amongst respondents. There was strong support from eleven respondents (R1-R4, R7-R12, R14) who reported that the ability to communicate effectively was a key non-technical skill needed by a DBA. Such communication spanned across a wide range of stakeholders within the organisation, including development and testing teams to senior management. Concerning communication skills, five respondents reported teamwork (R1, R4, R7, R8, R14), four respondents reported the need for social or interpersonal skills (R5, R6, R13) while three respondents reported the ability to act as a mentor (R4, R5, R13). These skills highlight that DBAs do not work in isolation. Respondents R3 and R9 spoke about the different contexts in which communication is important:

> ***R3:*** *"…communication is very important because a lot of the times people going to ask you okay what went wrong then you need to describe to them not in a technical DBA type of viewpoint but for their understanding…"* ***(31)***
> ***R9:*** *"…people skills … you need to be you need to be able to communicate your soft skills; you need to be able to communicate effectively with other people…"* ***(125)***

DBAs must continuously upskill themselves, and six respondents reported the need to be self-motivated (R1, R2, R4-R6, R13). Technology is continuously changing within the DBAs environment regarding new versions of software being released. Hence the DBA must ensure that they are always one step ahead. This also includes having the ability to problem-solve challenging issues they face or are consulted on by others as reported by four respondents (R1, R4, R6, R8). Respondent R1 spoke about problem-solving skills:

> ***R1:*** *"…I think you got to have very good problem-solving skills because a lot of the day is solving problems. (Hmm) … you got to be self-motivated…" (**5**)*

Time-management was also noted as an important non-technical skill required, not only for the role of the DBA but as an employee. Five respondents (R3, R5-R8) reported this as essential within the DBA role as there are many pressures within the working environment. Respondent R6 spoke about time-management:

> ***R6:*** *"… (hmm) but I would say as a corporate citizen, I mean you work in a corporate environment … time-management is a must for any person on that basis …"* ***(79)***

Table 5-2: Diverse range of non-technical skills category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1, 1.3, 1.4	Communication skills	1	1	1	1			1	1	1	1	1	2		1		**11**
1.3, 1.4	Self motivation	2	1		1	1	1							3			**6**
1.1, 1.4	Team work	1			1			1	1						1		**5**
1.4	Time-management skills			1		1	1	1	1								**5**
1.1, 1.2, 1.3, 1.4	Problem-solving	1			3		1		1								**4**
1.3, 1.4	Social skills					1				1				1	1		**4**
1.1, 1.2	Mentoring skills				1		1							1			**3**

5.2.3 Wide range of responsibilities category

With the DBA's diverse range of technical and non-technical skills as reported (section 5.2.1, 5.2.2), respondents also reported a wide variety of responsibilities which they perform daily. The majority of respondents reported that they spend a large proportion of their time performing activities related to the upkeep and safe-keeping of the databases under their administration, such as performance tuning and optimisation, ensuring availability, backup and recovery, general database administration, security, and database monitoring.

Ensuring that the databases are optimised regarding performance and availability to end-users was a high priority for respondents. This would extend to both internal end-users as well as external customers whose environments they might support. Strong support was shown by ten respondents (R2, R3, R5-R7, R9-R11, R13, R15) for DBAs to ensure that the databases which they administer, perform optimally and as efficiently as possible, such as optimising the database; and diagnosing poor-performing SQL. Respondent R3 spoke about performance tuning:

> ***R3:*** *"…and then the next thing they after I think is the performance tuning aspect … (hmm) … because we running such a lot of batch work in the*

background, especially at night, we need optimal performance to, because the batch window is about 7 hours long, by that time you should have everything done already – if there's any delay it'll have a major impact on the next day for the normal end users…" ***(27)***

Eight respondents (R1, R3, R5-R7, R12, R13, R15) reported availability as to ensure maximum uptime in all respects as the database is an important component of the business. End-users are reliant on the availability of the database to perform their duties and for organisations to conduct their business. Respondent R5 spoke about availability:

R5: *"I'd say to keep the database running for one, and running optimally so queries don't take too long … (hmm) … so that people can get their reports out … (hmm) … and making sure that everything is just running efficiently."* ***(61)***

The importance of availability of the data is intricately linked to backup and recovery whereby eight respondents (R3, R4, R7, R10-R13, R15) reported that the safe-keeping and integrity of the organisation's data and hence the ability to recover such data is a core responsibility of the DBA. Respondent R11 spoke about backups and other functions:

R11: *"… ja we do backups … our performance tuning … and then ... (hmm) … we do … (hmm) … like auditing as well on our databases …"* ***(148)***

Securing the organisation's data is important as the DBA is responsible for ensuring that the right users have access the data they need and included tasks such as auditing to ensure that users only access the data for which they were authorised. Five respondents (R8, R11-R14) reported on various tasks they perform in securing the organisation's data. Respondents R12 and R14 expressed the importance of security, given that in many instances DBAs must manage environments where the data is critical to the organisation's livelihood. Respondent R12 spoke about auditing:

R12: *"…if there's anything suspicious then I will raise a request for that and this is like for auditors purposes…"* ***(160)***

Many tasks related to the management of the database were broadly classified as general database administration, and this was reported by seven respondents (R3,

R6, R7, R9, R10, R13, R14). These included the setup and configuration of databases, and maintenance tasks to be performed which includes after-hours work and regular housekeeping tasks. These tasks are in support of critical goals such as maximum availability, and an optimally performing database. Respondent R13 spoke about after-hours work:

> ***R13:*** *"…and then after hours that's then maybe if we identify there's a … a release or a change that needs to happen that will impact the system or the ability of the system then we push those for after hours …"* ***(174)***

DBAs must continuously monitor the databases they support to ensure they are performing optimally, and this was supported by six respondents (R7, R9, R10, R12, R13, R15). This includes tasks such as monitoring storage, resource utilisation, performance and various health checks. Respondent R10 spoke about monitoring:

> ***R10:*** *"…we check for any file systems, tablespaces that have filled up … (hmm) … the previous night's backup that went through so that we know we can recover until the exact point-in-time..."* ***(134-135)***

DBAs must interact with and provide support to many stakeholders within the organisation as well as to external customers they support. Six respondents reported providing support to users within the organisation (R1, R3, R7, R10-R12) while providing support to external customers was reported by three respondents (R3, R10, R14). Support to internal end-users related to helping with various database related queries to provide the end-users with reports. End-user support was further highlighted by three respondents stating that they provide an advisory role to developers in the form of reviewing any database related code (R4, R11, R14). Respondent R14 spoke about end-user support:

> ***R14:*** *"…and we obviously you need that skill to vet the code and see if it's relevant and if something fail you can actually fix it that time especially before production however if it fail you follow the protocol, engage the developer, discuss it with the guys and ja …"* ***(188-189)***

Table 5-3: Wide range of responsibilities category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1, 1.2, 1.3, 1.4	Performance Tuning & Optimization		2	1		3	2	1		2	1	1		1		1	**10**
1.1, 1.2, 1.3	Availibility	3		2		1	1	3					4	2		3	**8**
1.1, 1.2	Backup and Recovery			1	1			2			1	1	2	1		3	**8**
1.2	General Database Administration			1			1	1		1	2			1	1		**7**
1.1, 1.2	Database Monitoring							1		1	1		2	3		1	**6**
1.2, 1.3	End-user Support	1		1				1			1	1	1				**6**
1.1, 1.2, 1.3, 1.4	Security						1		2			3	5	2	2		**6**
1.2, 1.3	Customer Support			2							3				6		**3**
1.2, 1.4	Developer Code Reviews				1							1			1		**3**

5.2.4 Automation of tasks category

Given the repetitive and potentially time-consuming nature of a large proportion of the DBA's work, nine respondents (R1, R3-R8, R14, R15) reported that they had automated many of the daily routine checks. This is to eliminate the need for manually performing the same checks daily as reported by respondent R15 who stated that manually performing such checks is a time-consuming process. In addition to having incorporated automation into their functions, eight respondents (R1, R7, R8, R10-R13, R15) also reported having a set daily routine which they follow hence they have reported having automated many of the routine tasks. Respondent R2 was the only respondent who explicitly stated that he has no set routines as each day is different. Respondent R4 spoke about automation:

> ***R4:*** *"…I'm really lazy in the respect that email or if I can run a couple of scripts in the background that saves me 25 minutes of mundane boring work, I'd do that."* ***(45)***

Table 5-4: Automation of tasks category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1, 1.2	Automated monitoring jobs	1		1	1	1	1	1	1						1	2	**9**
1.2	Set routines	1						1	1		1	1	1	1		1	**8**

5.2.5 DBAs have non-dba functions category

Feedback was divided about respondents having some form of responsibility or job function outside of the core responsibilities of what is required of a DBA. Five respondents (R5, R8, R12, R13, R14) reported that their sole function was to

perform activities related to the upkeep and maintenance of the database under their administration. Respondent R8 spoke about only performing database tasks:

> ***R8:*** *"…No non-DBA related functions – only other function that could be considered perhaps non-DBA is checking software strategy & compatibility of software against the database version…"* ***(116-117)***

Contrary to the above, four respondents (R1, R3, R4, R7) reported having responsibilities not generally considered being part of the DBA's role. This was emphasised by respondent R7 who was of the view that DBAs cannot solely specialise within the DBA space but needs to broaden their responsibilities. Non-dba functions included the administration and support of continuous integration systems used by development and testing teams and the installation of non-database related software. Respondent R7 spoke about having other responsibilities:

> ***R7:*** *"… I think that is something that is necessary as a DBA … as a DBA I think it's necessary that you cannot have like you know sort of the purest type of DBA …. purest? … maybe it's not the …"* ***(98)***

Table 5-5: DBAs have non-dba functions category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.2	No non-dba functions					1			1				1	1	1		**5**
1.2	Non-dba functions	1		1	2			2									**4**

5.2.6 Interface with a variety of stakeholders category

The requirement to interface with various departments, teams and stakeholders within the organisation was shared amongst the respondents interviewed. All the respondents except respondent R4 (R1-R3, R5-R15) reported that they had some requirement to interact or liaise with other departments within their organisation. This was either them providing a service, for example assisting developers, analysts or testers, or them having a dependency on other departments to function or complete tasks required of a DBA. For example, provisioning of storage and networking services from infrastructure departments. Respondent R13 spoke about interaction with other teams:

> ***R13:*** *"… so within the team we have like your teams that works with operating systems, storage, networks so when we work on the solutions those teams*

are integrated they work in the form of a project to deliver the solutions so the OS will build your operating systems, the storage will assign the storage and the network team will make sure your the servers you are able to access them…" ***(175)***

While respondents reported interaction with a variety of stakeholders within the organisation, there was particularly strong support for their working relationship with software development teams. Nine respondents (R2, R6, R7, R10-R15) explicitly reported that they have some form of interaction with or must work alongside software development terms. These interactions ranged from assisting with troubleshooting, fulfilling end-user requirements or providing guidance when it relates to the databases they support. Respondent R11 spoke about helping developers:

R11: *"… ja it is … because we deal with the … (hmm) … the application team … the project management team … (hmm) … the change management team on a daily basis … so …"* ***(149)***

Table 5-6: Interface with a variety of stakeholders category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.2, 1.4	Interact with other departments	2	2	1		2			1	1	4	2		3	2		**10**
1.2, 1.3	Interact with development teams		2				1	1			2	1	2	1	1	1	**9**

5.2.7 Working environment category

Findings reported that the DBA's working environment could be highly pressurised and stressful. There are many constraints within which DBAs must work. Six respondents (R3, R4, R7, R9, R10, R12) reported that they work within a pressurised environment due to the importance of the data they manage. This was emphasised by respondent R4 who said:

R4: *"…so there's a responsibility that you carry with you, and I find as a DBA in my perspective that's your biggest strength…"* ***(42)***.

Given the importance of the data administered in many production environments, five respondents (R2, R4, R5, R9, R12) reported that they work within a controlled environment with the additional pressure of financial implications as reported by

three respondents (R4, R6, R12) should anything go wrong or they fail to meet the service level agreement (SLA). Respondent R12 spoke about work pressure:

> ***R12:*** *"…I know that some of the services be it financial be it industrial be it whatever they have to report to a certain body and if they cannot provide the reports at a certain time they gonna be fined, there gonna be penalties so as a DBA I need to know that information so that I have to make sure that their system it's up and running…"* ***(162)***

It is therefore understandable that given these pressures, respondents also reported the need to be able to upskill in a short space of time as reported by three respondents (R3, R5, R15) in order to overcome issues, they might encounter.

Table 5-7: Working environment category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1, 1.2, 1.3, 1.4	Pressurised environment			1	1			2		1	2		1				**6**
1.1, 1.2, 1.3	Controlled environment		1		1	1				1			1				**5**
1.2, 1.3	Financial implications				1		1						1				**3**
1.3	Upskill in short space of time			2		1										1	**3**

5.2.8 Summary of the Skills and responsibilities (RDBMS environment) theme

DBAs currently working in an RDBMS environment presented a wide variety of skills and responsibilities required to perform as a DBA. The main findings are the following:

(i) Technical skills reported by respondents not only included the specialised skill set required to administer the database software, but also a diverse range of other technical skills and knowledge to understand the entire database solution and to provide support to other departments.

(ii) Non-technical skills presented strongly amongst respondents as being equally important as the technical skills.

(iii) Respondents have a wide range of responsibilities. Their primary focus is the upkeep and administering of the databases they support.

(iv) A small number of respondents reported that they also perform duties not generally considered to be the responsibility of a DBA.

(v) Many respondents have automated their daily routine checks to ensure the availability and integrity of the database systems.

(vi) Respondents reported that their role included interaction with a broad range of stakeholders within their organisations. This was either in a supporting

role, with software development teams featuring strongly, or having a dependency on infrastructure departments to perform their duties as a DBA.

(vii) The DBA's environment is a highly pressurised and controlled environment. Pressures include potential financial implications as well as the ability to upskill in a short space of time.

5.3 Theme - Background and education

The *"Background and education"* theme provides an insight into the background of respondents regarding their education and working experience. It reports on their transition from other roles they occupied into the current role as a DBA. Given this knowledge of their previous roles, it provides a better understanding of the diversity of skills that respondents have. Figure 5-2 below provides an overview of how this theme emerged.

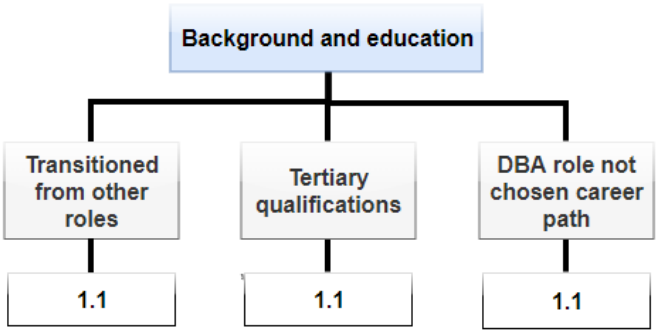


Figure 5-2: Background and education theme

5.3.1 Transitioned from other roles category

Respondents reported that they had experience within other roles before transitioning into the role of a DBA. Thirteen respondents (R1-R11, R13, R14) did not start their careers as DBAs. Respondents came from a diverse range of backgrounds, with eight of the thirteen respondents (R1-R8) having a background in software development. Respondent R5 spoke about his other roles:

> ***R5:*** *"Okay, … (hmm) …, I started off in programming really, … (hmm) … then moved to Systems Admin and more recently probably the last six years I've been doing DBA work…"* ***(60)***

Other notable careers included graduate level roles after graduating from university as reported by four respondents (R10, R11, R13, R14). It cannot be explicitly stated that these respondents had any specific role in mind after completing their tertiary qualification. Respondent R13 spoke about his job after graduating:

R13: *"… my background started when I completed … (hmm) … my studies … (hmm) … round about 2007 and then I joined the company as a junior IT specialist whereby in that time we just rotated across different sections you know just to learn about the operations in the organisation…"* ***(172)***

Findings showed that DBAs often have working experience in more than one role before becoming a DBA, and in certain instances transition across as many as four other roles as reported by respondent R14. Six other respondents (R2, R4, R5, R6, R7, R11) illustrated this point. Of these six respondents, respondent R6 had three prior roles while the other five respondents (R2, R4, R5, R7, R11) reported up to two prior roles. Respondent R14 spoke about his various roles in IT:

R14: *"… ok ja so I've fulfilled a couple of roles in IT sector … I think I started in IT 2000 … 2006 and 7 … 2006 going into 2007 ja … so I did mostly application support, network engineering, network administration during that time … (hmm) … later most to [CompanyNameWithheld] … [NameWithheld] … support satellite, virtual satellite support which also involves a lot of networking as well and later moved to … later moved to another company where I did some application support as well, moved to system engineering and then in 2012 moved to … (hmm) … moved to database management systems …"* ***(185-186)***

In contrast to the majority of respondents who reported that they did not start their IT careers within the role as of a DBA, only two respondents (R12, R15) only had working experience as DBAs.

Table 5-8: Transitioned from other roles category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1	Development background	3	2	2	1	1	2	1	2								**8**
1.1	Application / Desktop / Technical support										1	2		1	1		**4**
1.1	System administrator background		1			1	1			2							**4**

5.3.2 Tertiary qualifications category

Having a tertiary qualification was prominent amongst the respondents. Findings showed that thirteen respondents (R2, R3, R5-R15) reported having completed a tertiary qualification, ranging from a diploma to a degree, with respondents R12 and R14 citing that they both majored in two separate streams. Respondent R1 also

studied a tertiary qualification but reported that he did not complete his degree. Respondent R9 spoke about his tertiary qualification:

> ***R9:*** *"…Okay … so I started … (hmm) … 25 years ago I studied BCom at [UniversityNameWithheld] … I did a BCom Information Systems and then from there I worked my way up to System Admin..."* ***(122)***

The only exception to the above was respondent R4 who did not have a formal tertiary qualification but reported being self-taught. He further reported that he regrets not having studied more as he realises that it would have provided him with better opportunities. Respondent R4 said:

> ***R4:*** *"…I think if I studied a bit more, stayed a bit more current I could have probably commanded better, better salary … (hmm) … and possibly might have opened different opportunities like I'm seeing now in the BI and DSS space…"* ***(42-43)***

Table 5-9: Tertiary qualifications category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1	Tertiary qualification		1	1		1	1	1	1	1	1	1	1	1	1	1	**13**

5.3.3 DBA role not chosen career path category

Eight respondents (R4, R7, R8, R11-R15) reported that they did not start their IT careers with the intention of becoming a DBA. Three of the eight respondents (R11, R13, R15) reported that they were bursary students and were placed within the DBA role and hence did not have much choice as to the career they wanted to pursue. Respondent R12 reported that being a DBA was not her career of choice, but the need to be employed influenced her decision to take the role hence she studied both finance and IT. This was supported by respondent R15 who was strong in her response to the DBA role not being her chosen career path. She further explained that she is not passionate about the role and her main aim was to have the ability to earn a living. Respondent R15 said:

> ***R15:*** *"…No … that was not my intention … (hmm) … actually IT was not my intention … (giggles) … so basically I was studying to get money to be employable and survive …"* ***(202)***

Contrary, six respondents (R1, R3, R5, R9, R10, R14) reported that they have a passion for the DBA role citing various reasons. Respondent R14 purposefully transitioned to the DBA role because of an opportunity that arose but not with the intention of settling within the role for the rest of his career. He considers it as a stepping stone towards a management position. Respondent R1 differed to some degree and reported that it is a role that elicits professional respect and is aspired to. Respondent R1 spoke about why he transitioned:

> ***R1:*** *"…(Hmm) …, so I started off as a developer and when I was working as a developer, the DBA guys always seemed to be very knowledgeable and well paid, so I always thought it was a good position to aim for…"* ***(2)***

Table 5-10: DBA as a chosen career path category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1	Not chosen career				1			3	1			2	1	2	1	6	**8**
1.1	Purposefully transitioned	2		1		1				1	1				1		**6**

5.3.4 Summary of the Background and education theme

Respondents reported on their prior experience before taking up the role as a DBA including their educational background. The main findings are the following:

(i) The majority of respondents did not start their IT careers within the DBA role and transitioned from various roles. More than half of the respondents having a background in software development.

(ii) Almost all respondents have completed some form of tertiary qualification, ranging from a diploma to a degree.

(iii) More than half of the respondents reported being a DBA was not their career of choice. This was contrary to six of the respondents who were passionate about the role.

5.4 Theme - Training and upskilling

The *"Training and upskilling"* theme describes the various methods employed by DBAs to have the necessary skill set to perform as a DBA. This includes two primary types of qualifications that DBAs incorporate into their education, namely tertiary qualifications and vocational training received from database software vendors. While respondents considered both important to their educational requirements, it was evident that they had both positive and negative attitudes towards the benefits of each including the degree to which it is a requirement to perform as a DBA. This

was a recurrent theme across the research sub-questions. Figure 5-3 below provides an overview of how this theme emerged.

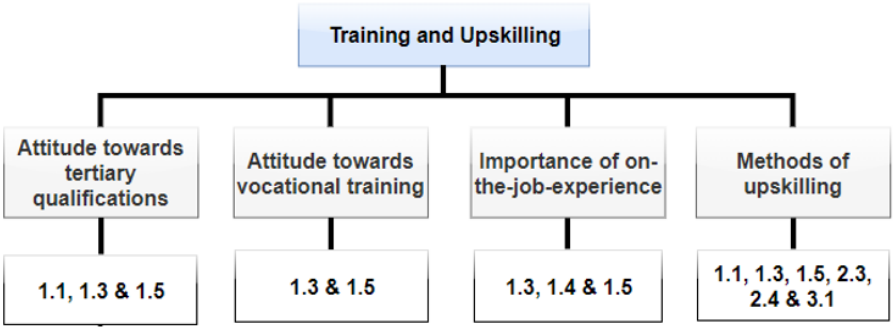


Figure 5-3: Training and upskilling theme

5.4.1 Attitude towards tertiary qualifications category

Opinion amongst respondents was divided on the requirement to have some formal tertiary qualification to perform as a DBA. Even though category *"Tertiary qualifications"* (section 5.3.2) reported that the majority respondents had completed some form of tertiary qualification, respondents felt strongly to distinguish between the benefits of having a tertiary qualification as compared to it being a requirement or enabler to becoming a DBA.

There was strong support and believe from respondents in that having a tertiary qualification is beneficial. Eight respondents (R3, R4, R8, R9, R11, R13-R15) showed positive support towards the benefit of having a tertiary qualification. This positive attitude did not indicate that a tertiary qualification is a requirement or enabler to perform as a DBA or that it provides DBAs with specific technical skills. Tertiary qualifications provided broader knowledge for the IT professional, and this was expressed by four respondents (R9, R13, R14, R15) who were similar in their views in this regard. Respondent R13 spoke about the benefits of a tertiary qualification:

> ***R13:*** *"…so I think the qualification prepares you to be able to be an independent thinker like you know like a person who is able to be given a problem and then you able to supply solutions yourself…"* ***(177-178)***

Eight respondents (R2, R3, R5, R6, R9, R11, R13, R15) reported that they did not consider it as being a requirement of the DBA role to have a formal tertiary qualification. This was further illustrated by three respondents (R2, R13, R15) who reported that there are very experienced DBAs within the industry with no formal

qualifications, yet they are very capable in performing their jobs. Other reasons included the need for a formal tertiary qualification is organisation dependent which was supported by three respondents (R3, R4, R14), specifying that corporates would more likely require you to have a formal tertiary qualification. Respondent R2 spoke about tertiary qualifications not being a job requirement:

> ***R2:*** *"…Okay, I work with a guy that has got 20 years of Oracle database experience with no qualifications - except matric and he's doing exceptionally good - he's our application production DBA…"* ***(16)***

Four respondents (R1, R7, R10, R12) differed slightly in their view and thought that DBAs should at least have some minimal form of qualification as part of the job requirement. Two respondents (R1, R12) reported that it is accepted within the industry that DBAs will have a tertiary qualification, and that many organisations consider it to be a minimum requirement. Respondent R10 expressed a strong opinion that having a tertiary qualification is important, even though he later reported to have not further pursued additional tertiary studies because he felt vocational training was more important. Respondent R10 spoke about the importance of a tertiary qualification:

> ***R10:*** *"… (hmm) … a formal qualification is very important within our field … (hmm) … I've done a national diploma in IT but I've decided not to do it further so that I could do my vendor training instead because I felt that it was a bit more important to … (hmm) … than going further on my studies because within Oracle there's a lot of courses say … (hmm) … with each technology you will have to go on a course …"* ***(139)***

Table 5-11: Attitude towards tertiary qualifications category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.1, 1.3, 1.5	Positive attitude towards tertiary qualification			1	3				2	1		2		2	2	1	**8**
1.5	Tertiary qualification not a job requirement		2	1		1	1			1		1		3		4	**8**
1.3, 1.5	Tertiary qualification as a job requirement	1						1			1		1				**4**
1.5	Tertiary qualification is company dependent			1	1										2		**3**

5.4.2 Attitude towards vocational training category

Findings showed that twelve respondents (R1-R3, R6-R10, R12-R15) expressed strong positive support towards the benefits of vocational training received from

database software vendors. This was further illustrated by eight respondents (R2, R3, R6, R8, R9, R11, R12, R14) who also included the certifications associated with such vocational training. Respondents believe that vocational training and being a certified DBA, not only enhances and adds technical skills in that it covers potential gaps within the DBA's knowledge, but also has other benefits such as validating and providing credibility to DBA skills. Findings showed that six respondents (R1, R6, R8, R13, R15) were also of the view that being a certified DBA has a positive influence regarding job opportunities and this was also indirectly reported by respondent R8 who stated that while certifications are not compulsory, it is encouraged by her organisation. Respondent R9 spoke about vendor training:

> ***R9:*** *"… puts you in a better position than when you have a degree because then you actually get into contact with the with the technical stuff immediately …"* ***(127)***

Even though respondents showed strong support towards the benefits of vocational training, four respondents (R3, R5, R13, R14) had some concerns with it being the sole means of gaining the needed skills as well as expressing some strong negative opinion towards it. While two of the four respondents (R13, R14) were also positive towards the contribution certifications made towards the DBA's skill set, they were concerned that without the ability to apply the knowledge gained, there was no benefit to being certified. The other two respondents (R3, R5) considered the scope of the courses to be too broad and as such, they were cramped for time. Respondent R13 spoke about the value of certifications:

> ***R13:*** *"… so I always say to the guys that go and write those exams and come and say I have the certification, I'll say that you have a certification but do you have the knowledge of how to use the tool itself and then most of the time it's a contrast you know in a sense that a person go get a certification but when comes to applying the knowledge …"* ***(178)***

Table 5-12: Attitude towards vocational training category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.3, 1.5	Positive attitude towards vocational training	2	2	1			2	4	1	1	3		3	1	2	3	**12**
1.5	Positive attitude towards certifications		1	2			1		1	1		1	1		2		**8**
1.3, 1.5	Certification influences job opportunities	1					1		1			2		1		1	**6**
1.3, 1.5	Negative attitude towards certifications			1		2								6	1		**4**

5.4.3 Importance of on-the-job experience category

Eight respondents (R1, R3-R5, R7, R8, R13, R15) reported in agreement that experience gained while working as a DBA was a key component to training and upskilling. Respondent R7 illustrated this by expressing a strong opinion towards the need for on-the-job experience rather than theoretical knowledge with respondent R1 agreeing with that sentiment as he noted that experience is preferred to training courses. Respondent R7 spoke about real-world experience:

> ***R7:*** *"…what has happened in real world you actually knocked your head, you actually fallen and you picked yourself up, you actually recovered from a disaster, you know those type of things … you (some words inaudible) migrated from one system to another, and those type of things so for me experience …"* ***(104)***

Table 5-13: Importance of on-the-job experience category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)														Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.3, 1.4, 1.5	On the job experience	1		2	1	1		3	1					1		1	**8**

5.4.4 Methods of upskilling category

Findings show that this category was recurrent across the entire interview for the majority of respondents. Respondents reported a wide range of methods they used to get the necessary skills to perform as a DBA. These methods of upskilling applied to both the RDBMS and big data environments.

A common view reported amongst ten respondents (R1, R3, R4, R7-R11, R14, R15) was for a DBA to engage in self-directed learning, hence being self-motivated to upskill and practice their existing skill set continuously. This type of learning was not limited to a single method to upskill or improve their current skills, but a large proportion of these respondents referred to them as reading as much as possible. Four respondents (R1, R2, R6, R7) further explicitly reported that as a DBA you need to be aware of emerging technologies while four respondents (R1, R4-R6) citing the internet as a source learning material such as online courses, reading white papers, user forums as examples of types of materials being sourced for learning purposes. Respondent R11 spoke about reading as part of the learning process:

> ***R11:*** *"You just need to read … that's all basically …"* ***(150)***

Another strong view with the majority of respondents was upskilling through training courses and becoming certified within the database vendor product suite which they supported. Six respondents (R3, R8-R11, R15) reported training courses as being an essential part of upskilling. This was further supported by five respondents (R1, R4, R9, R12, R15) who reported being a certified DBA as important. Respondent R9 was preparing for two certification examinations while he indicated that certification is encouraged by their organisation. Three respondents (R4, R10, R12) had strong views on putting the theoretical knowledge into practice by providing DBAs with a practice environment. Respondent R12 spoke about practical experience:

> ***R12:*** *"… besides the certification … besides going onto training … you need to do stuff hands-on so that ja … with no pressure because most of the time either if it's project related there will be timelines and there are costs involved which understand you end up being under pressure you know to deliver something that sometimes you don't even have knowledge on you just go and google stuff and you understand what I'm saying so I believe a lab environment would …"* ***(170)***

Three respondents (R3, R5, R11) also reported that more sponsored courses by the organisation would assist their ability to upskill. The lack of sponsored courses did not necessarily stop them from being able to upskill themselves given the broad range of methods and materials they use but would assist in their ability to upskill. Respondent R3 spoke about sponsored courses:

> ***R3:*** *"…I mean currently I'm doing also part-time SQL DBA work but I'm not certified yet but the company is not going to send me to do that but I don't think I will pay but maybe in the future but not now …."* ***(32)***

Table 5-14: Methods of upskilling category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1.3, 1.5, 2.4, 3.1	Self-directed learning	1		2	2			2	2	1	2	1			1	1	**10**
1.3, 2.3, 3.1	Training courses			1					3	2	2	2				1	**6**
1.3, 3.1	Certifications	2			2					1			1			1	**5**
1.3, 2.4	Awareness of emerging technologies	1	1				2	1									**4**
1.1, 1.3, 1.5, 2.3, 2.4	Internet-based content	1			3	3	2										**4**
1.3, 1.5, 2.1, 2.3, 2.4	Lack of company sponsored courses			4		1						1					**3**
1.3, 3.1	New vendor product releases												1	1		1	**3**
1.3, 2.4, 3.1	Practice environments				3						3		1				**3**

5.4.5 Summary of the Training and upskilling theme

"Training and upskilling" was a recurrent theme across the entire interview, signifying its importance to DBAs. The methods and requirements for training and upskilling as a DBA remained constant across both the RDBMS and big data environments. Respondents also expressed both positive and negative views towards the benefits of having a tertiary qualification as well as the need for vocational training. The main findings are the following:

(i) While the majority of respondents had a tertiary qualification, they did not consider it as a mandatory requirement. This did not prevent them from reporting on the benefits of being tertiary qualified.

(ii) Vocational training plays an important role in the DBA profession, given that the majority of respondents reported positively on this point. Its primary benefit relates to providing technical skills but it also has other benefits. A minority of respondents strongly felt that it must not be relied on as the only method, but it is important to practice the skill set obtained.

(iii) It is critically important to gain working experience as a DBA which was emphasised strongly by more than half of the respondents.

(iv) Strong support was shown for the need to engage in self-directed learning, being self-motivated to upskill continuously and to improve on their current skill set. Training courses and associated certifications also featured strongly.

(v) Support regarding sponsored courses by the organisation was reported to a lesser extent as respondents felt it could aid their ability to upskill should it be available for them to attend.

5.5 Theme - Adoption of big data

The "*Adoption of big data*" theme gives insight into the current exposure and experience relating to big data that the respondents had within their current organisations. These included an insight into their attitude towards both RDBMS and big data, as well as their view on their possible future involvement within the big data environment. Figure 5-4 below provides an overview of how this theme emerged.

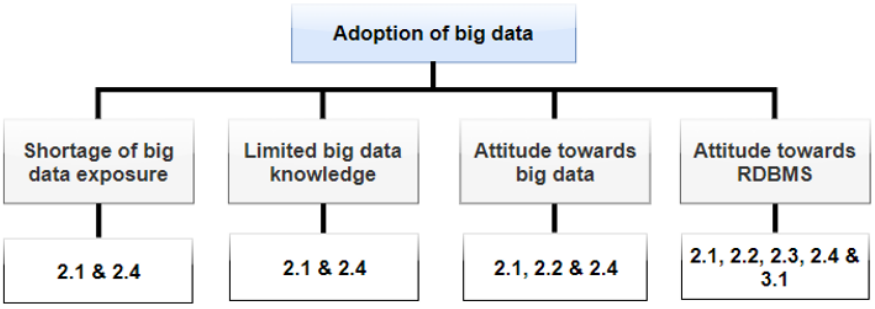


Figure 5-4: Adoption of big data theme

5.5.1 Shortage of big data exposure category

Findings showed that none of the respondents had any experience or exposure with big data in their current organisations. Nine respondents (R1-R4, R7, R8, R12, R13, R15) explicitly reported that they have no big data experience and or exposure while three of the respondents (R1, R2, R7) stated that they do not see any big data opportunities are available within their organisations. Respondent R1 further demonstrated this view by reporting that should he wish to gain exposure in big data then he would need to look for an opportunity in another organisation. One exception to the above was respondent R5, who reported that he is involved in big data to a degree in the light of him managing large volumes of data in data warehouses he is managing. Respondents R1 and R5 spoke about their involvement in big data respectively:

> ***R1:*** *"…I don't see myself going with what is now becoming known as big data unless I move to another company where, that's big data…"* ***(6)***
>
> ***R5:*** *"…Yes … I am involved in that to a degree at the moment (hmm) because we do a lot of archiving, we keep tons of data …"* ***(65)***

The findings regarding big data exposure extended further to respondents not having exposure to any big data tools. Four respondents (R1, R2, R9, R10) reported that they had some form of limited exposure to big data tools. Respondent R1 reported he had played around with Hadoop to become familiar with it while respondent R9 had some exposure with cloud technologies. Only two respondents (R2, R10)

indicated that they had a limited opportunity to do some research on Hadoop and non-relational database technologies such as NoSQL which was short-lived. Respondent R10 spoke about him experimenting with Hadoop:

> ***R10:*** *"… I've looked at … (hmm) … like I said I've looked at Hadoop … (hmm) … when there was a requirement last year …"* ***(143)***

Table 5-15: Big data exposure category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.1, 2.4	No big data exposure	1	3	1	2			2	1				1	1		1	**9**
2.4	Limited exposure with big data tools	1	1							1	2						**4**
2.1, 2.4	No big data opportunities in company	1	2					1									**3**

5.5.2 Limited big data knowledge category

While respondents did not receive exposure to big data within their organisations, they also reported that their knowledge on the area was limited. Seven respondents (R7-R9, R11-R14) reported that their knowledge on big data was limited, with four respondents (R7, R8, R11, R13) stating that they have read up on big data, but that is as far as they have gone. Respondent R8 stated that her knowledge of big data is only as far as reading up for the interview while respondent R12 explicitly did not want to comment at all on big data and reported that she only heard about it. Respondent R14 did report having received training in big data and referred to cloud computing technologies which is currently receiving much attention within their organisation. Respondent R8 spoke about her exposure to big data:

> ***R8:*** *"Read up to understand for the interview"* ***(119)***

Table 5-16: Limited big data knowledge category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.1, 2.4	Limited big data knowledge							2	2	2		1	1	1	1		**7**

5.5.3 Attitude towards big data category

Opinion regarding big data amongst respondents was somewhat divided. Even though respondents did not have any exposure to big data, some respondents

developed an interest in the area while others explicitly reported the lack of interest in wanting to pursue research in big data technologies.

Findings showed that six respondents (R1-R4, R7, R10) explicitly stated that they have no particular interest in big data technologies. These respondents all reported different reasons for the lack of interest in big data. There was particularly strong support in this regard from respondent R2 who wanted to continue practising as a DBA. He was particularly passionate about the DBA role within the RDBMS environment rather than big data and was steadfast in continuing it as a life-long career. Similar support was shown by respondent R7 who was interested in continuing his career in the RDBMS environment wanting to continue specialising and said that there is still a future for the RDBMS. Respondent R10 reported that given his limited interest in big data, he would nevertheless upskill if he is required to do so. Respondent R7 spoke about specialising in RDBMS:

> ***R7:*** *"… (hmm) … it is something that I would like to actually focus on … something that I would actually like to specialise on. RDBMS world or I mean non-Big Data sort of environments are not going to die yet, is not going to be dead yet."* ***(106)***

Respondent R1 wished to continue what he was doing and was in no haste to pursue big data. He would rather adopt a wait and see approach regarding the big data landscape before deciding what technology to pursue. Respondent R1 said:

> ***R1:*** *"…I wouldn't, well I haven't and I wouldn't to rush off to become a big data, to drop everything and learn all the big data toolsets – but definitely there are some toolsets that are standing out from the pack like Hadoop and Apache Spark and R I think is one of the other big ones – so there are definitely now becoming – it's becoming clear what the tools are that are going to move forward – so if I was moving that way then definitely there are things you can follow up on – that answer your question…"* ***(8).***

Six respondents (R3, R4, R7, R11, R13, R14) reported that they consider big data technologies as the way forward. Even though three of the six respondents (R3, R4, R7) reported that they have no interest in big data, they nonetheless acknowledged that it would be the prominent technology going forward. Respondent R3 said:

R3: *"…Well, ja, I've never worked in any big data environments so I think from my personal viewpoint that is definitely going to be the future…"* ***(33)***.

Six respondents (R3, R5, R10, R12, R13, R15) reported that they are not currently looking at or researching any big data technologies with respondent R3 previously indicated that this is due to his desire to pursue a career outside of IT.

Only three respondents (R6, R9, R11) showed positive interest in big data and suggested that they hoped to get involved with big data in the future. Respondent R9 indicated that while he is interested in big data, he does not have the time to do research and further his interest in big data as he would like to. Respondent R9 said: *"…it's fascinating to me I just wish I more time to go and look into it …"* ***(128)***. Respondent R6 said that while he is not looking into big data now, he is considering it in the future and had plans of his own to get into big data. Respondent R6 said:

R6: *"…Not yet, but it is something that I will be looking at … I mean I have some ideas of my own in terms of how to implement some of the solutions …"* ***(87)***

Table 5-17: Attitude towards big data category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.1, 2.2, 2.4	Big data technologies the way forward			1	1			2				1		1	1		**6**
2.1, 2.4	No interest in big data	1	2	1	5			2			1						**6**
2.4	Not researching big data technologies			1		1					1		1	1		1	**6**
2.1, 2.4	Interested in big data						1			3		1					**3**

5.5.4 Attitude towards RDBMS category

Findings show that seven respondents (R1, R4, R7-R10, R13) reported that the RDBMS environment would continue to exist even in the presence of big data. Respondent R1 reported that there is still value in the skill set of the traditional DBA and it is still important while respondent R9 reported that it would be a combination of both RDBMS and big data. Respondent R7 reported that even with the emergence of big data and its associated technologies, the RDBMS is not under threat and went further to report that there will still be a requirement for relational databases within smaller companies. Respondent R7 spoke about the continued existence of the RDBMS:

R7: *"…I would actually like to specialise on. RDBMS world or I mean non-Big Data sort of environments are not going to die yet, is not going to be dead yet…"* ***(106)***

Table 5-18: Attitude towards RDBMS category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.1, 2.2, 2.3, 2.4, 3.1	RDBMS will continue to exist	2			1			6	1	1	1			2			**7**

5.5.5 Summary of the Adoption of big data theme

The findings revealed the lack of big data experience and exposure amongst respondents. Respondents also had little desire to be involved in big data in the future. The main findings are the following:

(i) None of the respondents was currently working with big data in their current role of employment. The large majority of respondents also reported limited to no exposure to big data.

(ii) The majority of respondents reported little interest in big data with some respondents explicitly stating their lack of interest while others reported a continued interest in the RDBMS environment.

(iii) A few respondents reported a definite interest in being involved in big data in the future.

5.6 Theme - Skills and responsibilities (Big data environment)

The *"Skills and responsibilities (Big data environment)"* theme highlighted the respondents' opinions on the skill set requirements, and responsibilities required in a big data environment for the traditional DBA. Figure 5-5 below provides an overview of how this theme emerged.

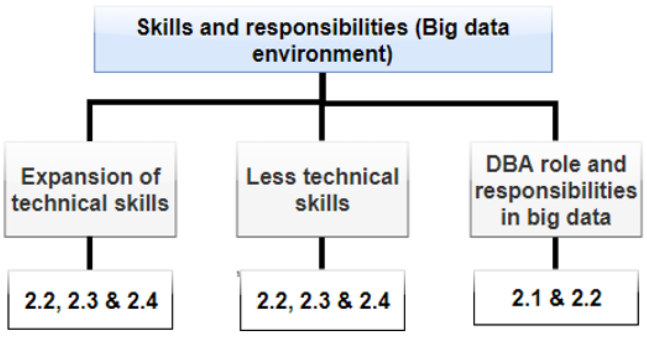


Figure 5-5: Skills and responsibilities (Big data environment) theme

5.6.1 Expansion of technical skills category

Findings showed that big data would not require existing DBA skills to be discarded but rather require DBAs to add skills or upskill in certain areas. Many of the skills respondents reported to being required overlap with the current skills required by traditional DBAs. Respondents reported that big data would be strongly focused on the management of the volumes and diversity of data in a big data environment as compared to the volumes and structured nature of the RDBMS environment.

The strongest codes supporting this category were around the skills required to be able to derive meaningful information and value from the volumes of data in a big data environment. These data-centric skills were specifically data analytics and data mining. Five respondents (R2, R3, R5, R6, R10) reported that DBAs would need to upskill in data analytics while three respondents (R2, R3, R14) reported the need to upskill in data mining. Respondent R10 reported that the focus would be more on analytics in big data as compared to what DBAs are traditionally accustomed to in the RDBMS environment. Respondent R2 supported upskilling in analytics and made a comparison with how the production DBA's role will change and further reported that it is possible that this role will no longer exist. Respondent R2 said:

> ***R2:*** *"…Yes of course, ja data mining … (hmm) … again analytical skills I think will be more important than what it is currently for production type DBAs…"* ***(20).***

Three respondents (R5, R7, R13) were more general about the skills required and reported that there would be a general awareness around data. Respondent R13 made an analogy in relation to programming and said:

> ***R13:*** *"…Coding definitely is … (hmm) … I think as much as we not coding the applications but we are coding the data..."* ***(181).***

Respondent R7 went further and reported that the focus would be around how the DBA can present the data and hence required an understanding of the data itself. Respondent R7 said:

> ***R7:*** *"…I think on the BI stuff, a Big Data DBA will you know sort of … I'm just going to use an example to provide reports on a daily basis of so much data … he needs to be able to actually understand how to get those reports …*

(hmm) … how to get the data … how to manipulate the data and how to present the data … so definitely yes…" ***(110)***

The adoption of cloud technologies within organisations is on the increase, and four respondents (R6, R11, R13, R14) reported that there would be a requirement to upskill in this area. Respondent R6 reported the requirement for skills related to cloud-based technologies to be able to accommodate the increased technologies associated with big data, while respondent R14 reported that there was an increasing adoption of cloud technologies within his organisation. Respondent R13 acknowledged the adoption of cloud technologies but reported some concern as to who would be administering the database environment in the backend. Respondent R14 spoke about the use of cloud:

R14: *"…Ja … at the moment we've been very very … (hmm) … working closely … basically it's cloud computing … it's the one that has been … (hmm) … the talk of our town because that's where the world is going basically … environment … you no longer going to be managing your own environment, we'll all be hosted in the cloud computing so that's the one basically that we've been looking at it closely ja … "* ***(196)***

Three respondents (R7, R10, R13) reported the requirement to upskill in programming to assist with the management of data. Respondents reported that there would be a slight change from the traditional DBA work performed in an RDBMS environment given that the focus will be more around converting the data into meaningful information. Respondent R7 was of the view that the need for programming skills would be an enhancement of the DBA's skills and should be considered when the DBA must work in a big data environment. Respondent R10 reported on the different methods of querying the database and said:

R10: *"… you would have to upskill the way you query the databases … (hmm) … even though you know the concept of the varying database …"* ***(143)***

Table 5-19: Expansion of technical skills category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.2, 2.3	Data Analytics		1	1		1	1				2						**5**
2.2, 2.4	Cloud technologies						1					1		1	1		**4**
2.3	Data Mining		1	1											1		**3**
2.3, 2.4	Focus on data					1		1						3			**3**
2.3	Programming skills							1			1			1			**3**

5.6.2 Less technical skills category

Contrary to the additional skills required as reported in category "*Expansion of technical skills"* (section 5.6.1), seven respondents (R3-R5, R10, R14, R15) reported that the DBA would require less technical skills in a big data environment. Even though four of the seven respondents (R3, R5, R10, R14) reported that DBAs would need to enhance their skill set, they still considered that there would be a requirement for less technical skills in a big data environemnt. These respondents reported that the increase in automated processes, and self-management features being introduced within DBMSs, will result in DBAs needing less technical skills. Respondent R15 referred to the introduction of "autonomous databases" hence her view was that DBAs would require less intervention and said:

> ***R15:*** *"…Most definitely yes … ja … because most of the things were manual and now things are automatic it will be just a matter of a click of a button …"* ***(210)***

Table 5-20: Less technical skills category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.2, 2.3, 2.4	Less intervention from DBA		4	1	2	1					1				2	2	**7**

5.6.3 DBA role and responsibilities in the big data category

Respondents were vague in reporting on what the role and responsibilities of a DBA in a big data environment will look like, and this produced a range of views. Seven respondents (R2, R5, R8, R9, R11, R12, R14) reported that they have no insight into what the role of a DBA would look like in a big data environment. Respondent R2 reported that it is possible that the DBA would be more business-oriented while the other five respondents (R8, R9, R11, R12, R14) were similarly short in their

response in reporting that they are not sure. Respondents R2 and R11 spoke of their uncertainty:

> ***R2:*** *"…I don't know, I've got no idea – maybe very boring … (hmm) … maybe that type of database administrator will also have to be more business type, user, or business person than a technical DBA…"* ***(18)***
> ***R11:*** *"…I'm not sure … I don't want to lie …"* ***(153)***

Contrary to the above, four respondents (R1, R3, R7, R10) reported that the role of the DBA in the big data environment would be similar to the current DBA role. They further reported different views as to what they consider the role will look like. Respondent R1 reported that the DBA role would be production focused, monitoring data feeds. Respondent R3 reported that DBAs would become more specialised. Respondent R7 reported that DBAs would be involved in more parts the system while respondent R10 reported DBAs would have similar routine checks and functions to perform. Respondent R10 spoke of the similarities:

> ***R10:*** *"… (hmm) … I would imagine it would be similar to what we do … they would also have to have backups … (hmm) … even though big data environments are highly redundant … (hmm) … I would also imagine it would … (hmm) … you'd have to check backups and all the things we do …"* ***(141)***

Three respondents (R1, R3, R7) reported that the responsibilities of DBAs would be the same. Respondents R1 and R3 had similar views and considered the role of the DBA in the big data environment to be similar to that of the traditional DBA. Respondent R1 later reported that he thinks there is a potential for a merger between the RDBMS and big data environments.

Table 5-21: DBA role and responsibilities in the big data category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.1	No insight into big data role		1			1			1	1		1	2		1		**7**
2.1	Similar role	1		1				1			1						**4**
2.1, 2.2	Responsibilities will be the same	1						1				1					**3**

5.6.4 Summary of the skills and responsibilities (big data environment) theme

Like the skills and responsibilities within an RDBMS environment, DBAs will require a range of technical skills and perform a variety of duties. Existing DBA skills will still be needed, and hence current skills will be enhanced to accommodate big data. The main findings are the following:

(i) The current DBA's skill set will be expanded and enhanced, and not discarded. Respondents considered that new skills that need to be added by DBAs would be data-centric. These include skills in data analytics, data mining, skills in cloud technologies and programming.

(ii) Contradicting the additional skills required, almost half of the respondents reported that DBAs would require less technical skills to manage the big data environment given that the technologies will be more self-managing.

(iii) Response to the responsibilities of the DBA in the big data environment was vague, and respondents either reported it as being similar or the same. Almost half of the respondents were unsure as to the role of the DBA in the big data environment.

5.7 Theme - Big data career pathing

The *"Big data career pathing"* theme highlights the current situation that is present within organisations regarding limited opportunities for DBAs to experience big data. It also provides an insight into the future career aspirations that respondents have. Figure 5-6 below provides an overview of how this theme emerged.

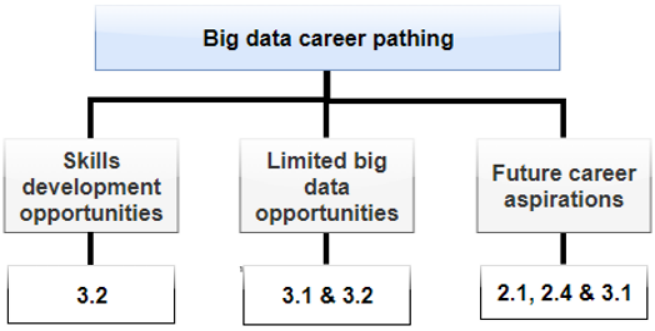


Figure 5-6: Big data career pathing theme

5.7.1 Skills development opportunities category

Findings show that the strongest code supporting this category was the view of four respondents (R1, R3, R5, R11) who reported that providing database vendor training courses will be an avenue to gain the needed skills. Respondent R5 spoke about sponsored courses:

R5: *"…Well I think they could … (hmm) … pay for courses and things like that, I mean that would definitely help … (hmm) … like more formalised courses … (hmm) … and Ja, that would help …"* ***(70)***

Contrary to the provision of courses as reported, three respondents (R9, R11, R12) had some concerns with only relying on training courses as a primary means of upskilling DBAs. Respondent R12 was primarily concerned with DBAs being allowed the opportunity to practice their skill set in a non-pressurised environment and not relying solely on the skills gained through training courses. Respondent R12 spoke about the ability to practice skills in a non-pressurised environment:

R12: *"… besides the certification … besides going onto training … you need to do stuff hands-on so that ja … with no pressure because most of the time either if it's project related there will be timelines and there are costs involved which understand you end up being under pressure you know to deliver something that sometimes you don't even have knowledge on you just go and google stuff and you understand what I'm saying so I believe a lab environment would …"* ***(170)***

A further two respondents (R9, R11) reported their concern with the cost of training courses. Respondent R11 also felt that the cost of such courses was a factor, but he was still in favour of them as a means of organisations enabling DBAs to upskill. In addition, respondent R9 preferred online courses given that training courses are costly in terms of both money and time. Respondent R9 spoke about online courses:

R9: *"… (hmm) … I'm thinking … in South Africa I specifically now I think they can still … they can profit to start using online training because a lot of the companies still send their people out you know to training centres or whatever … training environments … then they have to stay away from work for a week …"* ***(132)***

Table 5-22: Skills development opportunities category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
3.2	Company sponsored courses	1		1		2						1					**4**
3.2	Concern around training courses									4		1	2				**3**

5.7.2 Limited big data opportunities category

Findings show that there are either no plans within organisations to move towards a big data solution, or that they are not making it visible within their organisation to the extent that employees of aware of any such plans. Five respondents (R1, R2, R3, R5, R7) reported that there are no career pathing opportunities towards big data in their organisations. Four respondents (R1, R2, R3, R7) were of a similar view in that there is no career pathing towards big data and did not elaborate further. Respondent R2 spoke about the absence of career pathing:

> ***R2:*** *"…Okay, where we are you just a DBA and that's it – there's no career path …"* ***(22)****.*

Respondent R5 reported that there are no career pathing opportunities within his organisation because his view was that there is a perception that no other organisations are exploring big data, and as a result, organisations are not in a hurry to pursue this avenue. Respondent R5 said:

> ***R5:*** "…*(Hmm) …, well I can only speak for my organisation, I don't think they have really plans for that, I mean it's more kind of by default you know, nobody else is doing it so you know, I'll do it kind of thing…"* ***(69)***

Furthermore, seven respondents (R3, R5, R7, R9-R14) reported the lack of big data plans within their organisation. Of the seven respondents, four respondents (R9, R12-R14) reported that they are unaware of plans within their organisations to adopt big data while three respondents (R5, R10, R11) explicitly stated that there are no plans in this regard. Respondents R9 and R12 did not elaborate and just affirmed that they are unaware. Respondent R14 spoke about not being aware of big data plans:

> ***R14:*** *"…(hmm) … I don't have that information … ja … maybe [PersonNameWithheld] the one before me might have that information ja …"* ***(197)****.*

Respondents reported a variety of perspectives about the reasons as to the slow adoption of big data within organisations. Five respondents (R1, R4-R7) had slightly differing views as to the slow adoption of big data, but they all agreed in their sentiment that big data adoption is slow. Respondent R7 was strong of the opinion that there is a lack of understanding about the potential benefit that big data can

provide, and factors such as the potential cost of providing staff with the necessary skills and fearing change are prohibiting factors. Respondent R6 stated that organisations are realising the benefits of big data but adopting it slowly. Respondents R7 and R6 spoke about big data adoption:

> ***R7:*** *"… it's a longer term type of thing, they need to understand that you know right at the end that all the systems are going to be switched off unless we actually go to Big Data … (hmm) … prior to that it's the … it is sort of why do we need to go to Big Data … ? These are new technologies then we going to have to get like new you know sort of staff … you have to … (hmm) … redo all our infrastructure, those type of things …"* ***(113)***
>
> ***R6:*** *"…companies are realizing the importance thereof so they are moving towards it, maybe not as fast as they should…"* ***(88)***

Respondent R1 reported that there was a lack of advertised jobs, and this was his reasoning as to why he thinks the adoption of big data is slow. Three respondents (R10, R12, R15) suggested that it is dependent on whether there is a requirement in the organisation for a big data project hence they did not discard the possibility of being involved in big data in the future.

Table 5-23: Limited big data opportunities category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
3.1, 3.2	No career pathing towards big data	2	1	3		1		2									**5**
3.1	Slow industry adoption of big data	2			1	1	1	3									**5**
3.1	Not aware of company plans for big data									1			1	1	2		**4**
3.1	No company plans for big data					1					1	1					**3**
3.1	Upskill if required										1		1			1	**3**

5.7.3 Future career aspirations category

Eleven respondents (R1-R3, R5-R7, R10, R12-R15) explicitly reported that their future career aspirations did not reside within the big data environment and had various other career paths in mind. Eight of the eleven respondents (R3, R6, R7, R10, R12-R15) foresaw themselves within careers outside of both the DBA role as well as the big data environment, and these were various careers within IT. Respondent R3 spoke about pursuing another career:

R3: *"…If I'm still going to be in IT, I will, one of my things is rather to move more to the development environment, maybe adding a bit of other languages in, maybe a bit of architecture…"* ***(38)***

Six of the eleven respondents were more explicit in their career choices going forward. Three respondents (R3, R6, R14) reported moving towards an architectural role with respondent R14 saying: *"no … mostly more SQL architect …"* ***(197).*** A further three respondents (R1, R2, R5) wished to remain within the traditional DBA environment. Respondent R2 spoke about continuing as a traditional DBA:

R2: *"…(Hmm) … Not data big data as you've got it – I was actually looking at data warehouses…"* ***(21).***

Only two respondents (R9, R11) indicated that they foresee themselves working in a big data environment in the medium to long-term future. Respondent R8 indicated that there is a possibility should she be required to support a big data environment in future, that she will be open to the possibility.

Table 5-24: Future career aspirations category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.1, 3.1	Future outside of DBA role (other IT careers)			4			1	2			1		2	2	1	1	**8**
3.1	Architecture role			1			1								1		**3**
2.4, 3.1	RDBMS DBA	1	6			1											**3**

5.7.4 Summary of the Big data career pathing theme

There is currently very limited if any opportunities for DBAs to get exposure to big data within the organisations they are employed. The main findings are the following:

(i) Respondents considered training courses as an avenue to get the required skills for a big data environment, but there were a few concerns regarding cost and time associated with these training courses.

(ii) There is very limited or no big data activity within the organisations in which respondents are employed.

(iii) The majority of respondents want to pursue other IT careers in future.

5.8 Theme - Transitioning to a big data role

The *"Transitioning to a big data role"* theme describes the respondents view as to the readiness of traditional DBAs to transition into a role within the big data environment

and the re-skilling requirements to enable such a transition. Figure 5-7 below provides an overview of how this theme emerged.

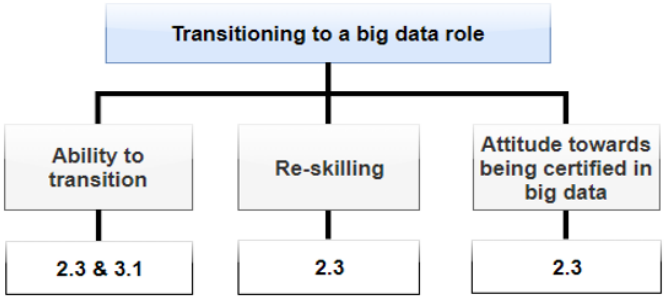


Figure 5-7: Transitioning to a big data role theme

5.8.1 Ability to transition category

Eight respondents (R1, R4-R6, R9-R11, R15) reported positively about the ability of DBAs working in an RDBMS environment to transition to a role within big data. Respondents reported various reasons for them believing that DBAs have the skills to transition. Two respondents (R1, R11) reported that they do not think that there's any difference between the roles while another two respondents (R5, R10) considered the roles to be similar. Respondent R15 reported that DBAs would have less manual work to perform as many of the processes will be automated. Respondent R15 spoke about the automated environment:

> ***R15:*** *"…Most definitely yes … ja … because most of the things were manual and now things are automatic it will be just a matter of a click of a button …"* ***(210)***

Two respondents (R2, R14) were less committed in their response and said it would be dependent on experience and the tools used, respectively.

Table 5-25: Ability to transition category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.3, 3.1	DBA has skills to transition	4			1	1	1			1	2	2				2	**8**

5.8.2 Re-skilling category

Findings show that respondents had a range of views about the need for DBAs to re-skill to enable them to work in a big data environment. Seven respondents (R2-R6, R13, R15) reported that there would be a requirement to re-skill to work in a big data

environment. All seven respondents reported different reasons as to why they thought it is a requirement for the DBA to re-skill. Of the views expressed, respondent R6 reported that there would be a branching of different specialists while respondent R15 was less committed to explaining her reasoning. Respondent R6 spoke about re-skilling:

> ***R6:*** *"…there would be branching into many different aspects and there would definitely be re-skilling required…"* ***(85)***

Three respondents (R7, R11, R14) explicitly reported that they do not see the DBA requiring any formal re-skilling. Respondent R11 and R14 were short in their response to the question of re-skilling and did not elaborate. Respondent R7 reported that he did not consider the DBA's role to be under threat and did not see the need to immediate re-skill. Respondent R7 spoke about not re-skilling:

> ***R7****: "…Yes, I feel that … well not formally … I don't think that the the the traditional DBA … (hmm) … sort of … (hmm) … role is that much of a threat yet, whereby you do have go learn something else right now, however it will get there…"* ***(108)***

Table 5-26: Re-skilling category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.3	Re-skilling required		1	1	3	1	1							2		1	**7**
2.3	No re-skilling required							1				1			2		**3**

5.8.3 Attitude towards being certified in the big data category

Respondents attitude towards being certified was to a large extent positive. Eight respondents (R1-R3, R9-R12, R14) had a positive attitude towards getting certified in big data with four of the eight respondents (R2, R3, R9, R11) showing strong support for certification being essential. Respondent R2 was short in his affirmation towards certification and said: *"…Certification obviously…"* ***(19)****,* while respondent R9 was also confident in the requirement to be certified in big data. Respondent R9 said:

> ***R9:*** *"…I think you'll definitely have to re-certify … ja … but also some of the … (hmm) … some of the … (hmm) … business intelligence stuff … (hmm) … is*

part of the DBA track so you'll probably be able to use a lot of that already … (hmm) …" ***(129)***

In addition, three respondents (R7, R10, R14) considered the certifications within big data as add-ons to existing certifications. Their views were that it would be an extension of the current certifications that DBAs have within the RDBMS environment.

Contrary to the positive response from respondents, three respondents (R4, R13, R15) were less optimistic about the benefits that certifications hold in terms of being able to perform as a DBA. Respondents R4 and R15 were more in favour of having on-the-job experience rather than being certified, even though respondent R4 later reported that there would be a requirement for a minimal type of qualification in the job market. Respondent R13 reported that he could not identify what benefit it holds. Respondent R15 spoke about the need for actual working experience:

R15: *"… because you write a certification first without … without even worked on that thing … then after the certification you start working on the thing … experiences the thing so according to me it should be the other way around … you should first work on the thing and then just get a certificate …"* ***(210)***

Two respondents (R5, R9) were unsure of its benefits.

Table 5-27: Attitudes towards being certified in the big data category

Interview questions	Strongest codes in category	Respondents from which codes emerged (R1-R15)															Strength of code (N)
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2.3	Certification in big data is essential		3	1						1		1					**4**
2.3	Positive towards getting certified in big data	3									1		1		1		**4**
2.3	Big data certifications will be add-ons							2			1				1		**3**
2.3	Negative towards getting certified in big data				2									2		6	**3**

5.8.4 Summary of the Transitioning to a big data role theme

Respondents reported with confidence in the ability of the DBA working in the RDBMS environment to transition into a role within the big data environment. The main findings are the following:

(i) The majority of respondents reported positively on the DBA's ability to transition into a role within the big data environment successfully.

(ii) Almost half of the respondents reported the need for DBAs to re-skill in big data. A few respondents explicitly reported that no re-skilling would be needed.

(iii) More than half of the respondents had a positive attitude towards being certified in big data, while some respondents further illustrating it will be in addition to RDBMS certifications. A few respondents were less optimistic, with two respondents preferring experience over certification.

5.9 Summary of Chapter Five

Chapter Five presented the findings relating to the difference in skills and responsibilities of DBAs working in an RDBMS environment as compared to a similar position within a big data environment. The findings of seven themes and the underlying categories were presented in detail.

The findings showed that DBAs have a wide variety of skills and responsibilities, and the big data environment will require an enhanced set of these skills and responsibilities. Current DBA skills will not be discarded but enhanced. Respondents in this study however reported a lack of exposure and experience regarding big data and this is attributed to their organisations not exploring big data at present. While respondents are confident in the DBA's ability to transition to a role within big data, the majority of respondents are considering furthering their careers within other roles in IT. Only a minority of respondents foresaw themselves within careers in the big data environment.

A detailed discussion of the seven themes reported in this chapter will be presented in Chapter Six.

CHAPTER SIX: DISCUSSION OF FINDINGS

6.1 Introduction

Chapter Five presented the findings from the 15 interviews conducted with individuals within the DBA role. These findings were presented in relation to the seven themes which emerged. The themes are as follows:

(i) Skills and responsibilities (RDBMS environment).
(ii) Background and education.
(iii) Training and upskilling.
(iv) Adoption of big data.
(v) Skills and responsibilities (Big data environment).
(vi) Big data career pathing.
(vii) Transitioning to a big data role.

Restatement of the research problem to be addressed in this study:

There is currently a shortage of specialists to administer big data environments in organisations. Even though DBAs manage the organisations traditional RDBMS, they do not have the required skills needed to work in a big data environment. Organisations might therefore not be able to extract the potential value that big data holds if they do not have the necessary skilled staff to manage and analyse big data within the organisation.

This study is attempting to answer the following main research question:

1. What is the skills gap between the DBA working in an RDBMS environment and a similar role within a big data environment?

The research sub-questions underlying the main research question are:

1. What are the current skill set and responsibilities of a DBA working in an RDBMS environment?
2. What are the skills and technology requirements of a big data environment?
3. What career plans do organisations have for DBAs with regards to big data management?

6.2 Theme - Skills and responsibilities (RDBMS environment)

The *"Skills and responsibilities (RDBMS environment)"* theme relates to the skill set and the responsibilities of the DBA working in an RDBMS environment to perform their daily function. (A summary of the findings for this theme can be found in section 5.2.8).

The discussion of this theme follows two threads:

(i) Skills of the DBA (RDBMS environment).

(ii) Responsibilities of the DBA (RDBMS environment).

6.2.1 Skills of the DBA (RDBMS environment)

The findings suggest that the DBA role requires a well-rounded, technically skilled individual, relating to both technical and non-technical skills. Technical skills include being proficient in all aspects relating to the management of the database software for the supported vendors, but importantly also incorporates broader knowledge of the entire database solution. This includes having knowledge of the underlying infrastructure supporting the database as well as the surrounding systems that interact with and have a dependency on the database. Non-technical skills are no less important given that the role requires interaction with many different stakeholders across the organisation.

The role fulfilled by the DBA within organisations should not be underestimated because the database in many instances forms an integral part of the organisation's business and livelihood. Unexpected or extended periods of downtime can have potential financial implications, and it is thus important to have an individual who is technically capable of ensuring minimal downtime is experienced (section 2.3). Three respondents in this study reported the potential financial implications if certain processes are not performed timeously (section 5.2.7).

Even though it is critical that DBAs are highly proficient in the database vendor software they support, it is important to incorporate knowledge of other areas such as the underlying infrastructure supporting the database (section 2.3.1, 5.2.1). DBAs must often troubleshoot performance related issues, and this is not always related to the database. Other factors such as slow disk input-output (I/O), network latency, faulty or corrupt disks can all contribute to an un-optimised database or potentially lead to an outage. Therefore, being knowledgeable about infrastructure related components supporting the database such as operating system, networking, storage and potentially hardware does not imply that the DBA must resolve such issues, but

it helps the DBA in identifying potential bottlenecks and issues. It is important that the DBA can communicate on a technical level with more specialised teams such as system administrators, and those responsible for supporting the infrastructure.

The requirement to have programming skills may not be directly related to the role of the DBA even though it is considered as part of the skill set of a DBA (section 2.3.1). This study has reported on two potential explanations for this.

- Firstly, respondents' interface with a variety of stakeholders within their organisations and there was strong support for the working relationship with software developers amongst other departments. Many of their functions include assisting software developers in troubleshooting performance related issues and understanding how the application interacts with the database. Respondents in this study have also transitioned from a variety of different roles before fulfilling the role of a DBA, and there was strong support for having a background in software development (section 5.2.6, 5.3.1).
- In addition to programming skills, respondents also reported other development tools such as SQL and scripting. DBAs need to be proficient in SQL as this is the language used by relational databases to query its data (section 2.2.1). There is a potential dependency on DBAs for their knowledge of SQL by software developers to either aid with troubleshooting errors or optimising SQL (section 2.3.1). Programming is further extended towards the requirement for scripting skills to perform various functions. This includes automated monitoring scripts to ensure the wellbeing of the database such as ensuring that storage on the file system is adequate, checking for errors in the database logs, monitoring backups to ensure that minimal downtime is experienced.

As such, it is possible to see the various interactions DBAs have across the organisation, both from a supporting role for software developers, software testers and various business units, to having a dependency on other departments such as different teams responsible for infrastructure. This extensive skill set could also be ascribed to many organisations employing database solutions from multiple vendors (section 2.3.1) (Holt *et al.*, 2015). DBAs thus must extend their skill set across multiple vendor products (King, 2015). Different vendors use different architectures and platforms upon which their database solutions are deployed. These include deploying their software on single or multiple operating systems such as MS Windows and variations of Unix and Linux.

This study has also found that respondents place great importance on non-technical skills. These skills are largely related to skills required to interact with other individuals such as skills relating to communication, teamwork, social and mentoring (section 5.2.2). These skills are related in that DBAs do not work in isolation and the importance of these skills can be understood given that the DBAs interact with many different stakeholders within the organisation. This is emphasised by Bailey (2014) who states that DBAs work within a team and requires various social skills to interact and work as a team. This includes skills such as multi-tasking, problem-solving and stress management. This study also emphasised the importance of the DBA's ability to communicate well because they must communicate across many levels of the organisation, including senior management (section 5.2.2). The ability to troubleshoot and problem-solve also requires the DBA to be self-motivated in researching where the potential bottlenecks in the system can be. As previously discussed, the DBA's technical and non-technical ability is needed across the entire database solution incorporating both humans and technology.

6.2.2 Responsibilities of the DBA (RDBMS environment)

The need for a DBA to be technically and non-technically skilled can be closely attributed to the broad range of responsibilities and functions they perform. The role not only focuses on the management of the database software, but findings suggest that in certain instances, the DBA must also incorporate non-dba related functions. As a result of the extent and potentially repetitive nature of their duties, DBAs have automated many of their tasks. This further extends to the requirement to provide support to a wide range of stakeholders within the organisation.

This study found that even though the management of the database software was reported as the primary function of the respondents, being a DBA involves a wide range of activities and responsibilities to ensure the database functions as required. In general, the DBA role includes performing tasks such as database design, software installations, the configuration of the database, upgrades to newer versions, general database administration, monitoring and maintenance, security, governance, and regulatory compliance as well as backup and recovery (section 2.3.2). Findings of this study reported that the functions performed by respondents in this study are consistent with many of the functions outlined in the literature (section 2.3.2). While respondents in this study showed strong support for the various functions related to the maintenance and upkeep of existing databases they administer (section 5.2.3), none of the respondents explicitly reported being involved in the design and

development of the database as outlined as part of the functions performed by DBAs (section 2.3.2).

Respondents in this study show strong support for the various functions related to the efficient performance of the databases they administer as well as the activities relating to the safe-keeping of the organisation's data. This included performance and optimisation, the availability of the database as well as backup and recovery functions to ensure the integrity and security of the data under their administration (section 5.2.3). The importance of the data can be related to both databases supporting the production and development environments within the organisation. Organisations place great value on the data they have in their possession as this is reflected in the findings of the study and the literature (section 5.2.7, 2.3). Respondents in the study reported on potential financial implications should they breach SLAs, but this can also be associated with not being able to deliver software. Organisations are also making use of shorter, agile development sprints to deliver software and features quicker to market (Holt *et al.*, 2015), and this will also place increased importance on the dependability and availability of the database that supports these development efforts. Having to rebuild a database, or potentially not being able to recover data, even in the development environment can place a burden on delivering either a service to the organisation's customers or bringing a product or software update to the market on time.

Some of these tasks reported in the study can be repetitive and time-consuming, such as continuously monitoring the health of the database, and this could suggest why many of the respondents have resorted to automating such mundane yet important functions (section 5.2.4). It is possible to understand that given the critical importance of the database in many organisations, the DBA needs automated processes to continuously monitor the database and inform them of potential issues or as soon as an issue occurs (section 2.3.2).

While the DBA role is considered a specialist role (section 2.3), findings in this study reported that DBAs also have non-dba related work they are required to perform (section 5.2.5). There is some evidence to suggest that given the broad technical skill set presented by respondents, as well as having transitioned in many instances from other roles, DBAs can adapt and perform functions considered outside of the scope of the DBA role. In addition to managing the database, there are other supporting functions associated with their role. DBAs interact with many stakeholders within the organisation, but this study showed a strong working

relationship with software developers (section 2.3.2, 5.2.6). DBAs often must guide developers when troubleshooting SQL or reviewing code that is to reside on the database (section 2.3.2, 5.2.3). This is not a one-way interaction whereby the DBA provides a service to other departments but also in other instances whereby the DBA is dependent on others to perform their duties (section 5.2.6).

6.2.3 Summary of the Skills and responsibilities (RDBMS environment) theme

The DBA role is specialised and complex. Individuals fulfilling this position need to be skilled in both a diverse range of technical as well as non-technical skills. It is plausible that this diverse skill set is closely related to the many duties needed to be performed within the RDBMS environment including skills inherited from occupying other roles. It is not a role fulfilled in isolation though, but a symbiotic relationship exists between the DBA and many stakeholders within the organisation whereby the DBA extends their expertise and services to others but is also dependent to an extent on the services of other departments to fulfil their duties.

6.3 Theme - Background and education

The *"Background and education"* theme provides an insight into the composition of individuals within the DBA role concerning prior roles occupied and their educational background. This theme is linked to the *"Skills and responsibilities (RDBMS environment)"* theme (section 6.2.1) in the context of the skill set presented by respondents. (A summary of the findings for this theme can be found in section 5.3.4).

There is evidence to suggest that a tertiary qualification is important to have but this evidence does not suggest that a tertiary qualification was completed with the intention of pursuing a career as a DBA. It is further possible to suggest that the DBA role is not a chosen career from the outset, but it is a position aspired towards or considered because of experience gained in other roles within IT. This is primarily supported by more than half of respondents having experience within software development.

While it is acknowledged that a DBA would generally have some form of tertiary qualification and that it can be advantageous regarding job opportunities (section 2.3.3), the findings of this study suggest that having a tertiary qualification is not necessarily as a result of wanting to pursue a career as a DBA (section 5.3.1, 5.3.3). However, it is possible to suggest that most often an individual fulfilling the role of a DBA would have some form of tertiary qualification, either a diploma or degree, but

in a more generalised field such CS or a stream within IT (section 2.3.3) (McGehee, 2009:51-52). This is highlighted by the majority of respondents having transitioned to the role rather than pursuing it from the outset (section 5.3.1, 5.3.3). Completing the tertiary qualification thus enabled them to be considered a suitable candidate in the IT job market or wanting to pursue more senior positions (section 2.3.3). One respondent (R14) reported that he wishes to pursue a managerial role in future and that the role of a DBA provides valuable input in his pursuit thereof, while another respondent reported the need to simply be employed (section 5.3.3). This to be consistent with the DBA profession not commonly being a field considered amongst individuals given the limited availability of tertiary institutions providing degrees in database administration and that becoming a DBA is often "through a set of often-random circumstances" (McGehee, 2009:52).

The diverse technical and non-technical skills presented by respondents in this study can also be attributed to the experience gained in other roles within IT. There is no clear evidence to suggest that respondents had any intention from the outset to become a DBA. More than half of the respondents reported that being a DBA is not their chosen career path (section 5.3.3). Because data is important to the organisation, DBAs often fulfil a critical role within the organisation. It can therefore be considered that it is important for DBAs to be confident in their technical skills and ability, and thus individuals more often transition from other technical roles into the DBA role rather than specialising as a DBA from the start of their career. Almost half of the respondents have purposefully transitioned to the DBA role after having had exposure within other fields within IT. Some of the respondents have experience across multiple different technical roles (section 5.3.1). This can be further supported in that organisations often promote individuals into the DBA role having experience within other technical roles (McGehee, 2009:56).

6.3.1 Summary of the Background and education theme

The DBA role is a position which is often not chosen from the outset, but individuals transition towards the role after gaining experience in other technical roles within IT. This was supported by the findings which reported that more than half of the respondents reported it was not their intention from the outset to become a DBA. Often it is possible that DBAs are individuals with a diverse range of technical skills inherited from working experience in other fields within IT and that it is a progression towards the DBA role. This can be further explained by the fact that most DBAs will have some form of tertiary qualification not because of it being a mandatory

requirement of the DBA role but because they pursued other careers before transitioning into the DBA role.

6.4 Theme - Training and upskilling

The *"Training and upskilling"* theme provides an insight into the diversity of methods employed by DBAs to learn new skills as well as enhance their current skill set. This is a further extension to the *"Background and education"* theme (section 6.3) in relation to what DBAs consider to be an important avenue in gaining the technical skills and knowledge required in the DBA profession. (A summary of the findings for this theme can be found in section 5.4.5).

6.4.1 Wide range of upskilling methods

Findings suggest that upskilling and learning new skills is a continuous process utilising a range of methods. There is evidence to suggest that DBAs are primarily responsible for their own upskilling while also relying on technical skills acquired from vendor training courses if such courses are sponsored by their organisations. On-the-job experience is considered an essential part of a DBA's training hence the theoretical knowledge gained through their own research and training courses cannot solely be relied upon.

DBAs spend a considerable amount of time and invest in a wide range of methods to acquire new skills and to enhance their current skill set (section 5.4.4). The role requires continuous learning and enhancement of skills to ensure that DBAs are technically proficient in managing and maintaining the databases they administer given the critical importance of the database to many organisations (section 2.3).

As with many careers, there are existing guidelines and methods to learn and develop the skill set required by the role. In the case of the DBA, McGehee (2009:59-61) describes the wide variety of methods employed by DBAs which include formal classroom training, seminars and workshops, conferences, user group meetings, online training, self-study book learning, magazines and other publications, websites, blogs, news, RSS feeds, and on-the-job training opportunities.

In this study two-thirds of the respondents relied to a very large extent on their own ability to investigate and research technologies and skills they require to perform their duties, while almost two-thirds also reported on the training and certifications received from the database software vendors (section 5.4.4). Adriana and Holanda

(2016), who surveyed 44 DBAs in Brasilia in 2015 and 2016, found that respondents employed a variety of methods to get skills required for the job. It included self-study, reading articles, forums, vendor training and to an extent, conferences, and workshops. Of these respondents surveyed, more than 95% of the respondents incorporated what was termed “self-study” into their training while training received from database software vendors was evident in only approximately half of the respondents.

Several reasons could explain why DBAs have incorporated self-study into their training to the extent to which they have. While training provided by the database software vendors is a highly effective means to learn the technical skills required of a DBA, it is often expensive and time-consuming (McGehee, 2009:59). These courses are often conducted outside of the organisation at either the vendor’s premises or training centre. The opportunity to attend such courses are also not always being promoted by all organisations as was the case in a study conducted by Holt *et al.* (2015) who surveyed various roles responsible for database management. While no reasons for limited training was provided in the Holt *et al.* (2015:179) study, they found that “23% of respondents were never given the opportunity to undertake formal training” and “28% were never encouraged to obtain certification”. As such, DBAs will not always be afforded the opportunity to attend such courses on a regular basis for every new database version or when new features are being released. Many organisations also make use of database solutions from multiple vendors and in many instances, DBAs must also support multiple vendor products, making it costly for organisations to continually upskill their employees. Limited opportunities to attend vendor training courses was reported to a lesser degree in this study whereby three respondents would like to receive more training opportunities through vendor courses. DBAs will also often find themselves in critical situations needing them to either react quickly or find solutions to problems they have not encountered before. It is often the case that they will search the internet, making use of online manuals from the database vendor or various websites or online user forums to assist in resolving the issue.

Apart from all the methods employed by DBAs to upskill themselves, it is important to understand that accompanying the newly acquired skills and knowledge with actual working experience is of utmost importance (section 5.4.3). Even though it is evident that many DBAs direct their own learning and make use of training received by database software vendors, just over half of the respondents found it necessary to emphasise that all the knowledge gained is of little benefit to them if they do not

experience real-life situations or having the ability to practice their skills in a non-pressurised environment. Polakowski (2009) also considers vendor certification programs of great benefit to DBAs but only if those skills acquired are exercised in real-life situations.

6.4.2 Attitudes towards tertiary qualifications and vocational training

Findings suggest that vocational training received from the database software vendors is more beneficial when wanting to add or enhance the technical skills of a DBA. Having a tertiary qualification is advantageous, but it is not a requirement or necessarily an enabler to perform as a DBA.

As previously discussed in section 6.3, almost all of the respondents in this study have a tertiary qualification and they considered it to be very beneficial. However, their attitude towards a tertiary qualification being a job requirement showed the contrary (section 5.4.1). Just over half of the respondents in this study reported that it is not a job requirement, with two respondents citing that they are aware of experienced DBAs with no tertiary qualifications. Another two respondents in this study also reported not to have completed a tertiary qualification. This lends support to the researcher's view that the DBA role can be fulfilled without a having tertiary qualification. It is accepted that a DBA would have some form of tertiary qualification, but it is not a mandatory requirement (section 2.3.3). It is however beneficial to them in that employers are in favour of individuals who have a degree related to either CS or IT (Moore, 2010).

The DBA role requires an individual who is technically skilled. This suggests that a tertiary qualification does not necessarily provide any such technical skill set which directly relates to the management of the database. There are a few possible explanations for their attitude towards tertiary qualifications. One possible reason might be that because their tertiary qualifications were not directly related to the field of database administration, that they were unable to ascribe any direct benefits to the work they are currently performing as DBAs. This can also be considered in the light of the availability of university degrees, specialising in database administration is not that common (McGehee, 2009:51).

Findings in this study suggest that DBAs consider training and associated certifications by database software vendors to be more important than having a tertiary qualification. Certification programs provided by database vendors such as Oracle, Microsoft, and IBM are common amongst DBAs in order to get upskilled

(Tittel, 2003). This was further illustrated in a survey of 44 DBAs in Brasilia across 2015 and 2016 whereby almost half of the DBAs used the vendor training as a means of training and hence enhancing their skill set (Adriana & Holanda, 2016). It is possible to suggest that because the DBA role is considered a specialist role (section 2.3), DBAs feel confident in getting the necessary technical skills from the database software vendors (section 2.3.4). This confidence is in the light of certification exams associated with these training programs (section 2.3.4). Many respondents reported additional benefits such as providing credibility to their skill set, validating what they currently know, filling gaps in the knowledge and influencing job opportunities (section 5.4.2). This would suggest that DBAs place a certain degree of confidence and accomplishment in their ability and expertise as a DBA once they have undergone such training and or have been certified in the database vendor software.

6.4.3 Summary of the Training and upskilling theme

The DBA role is complex and diverse regarding the skill set required and functions to perform as previously discussed (section 6.2). DBAs are often required to administrator databases from multiple database vendors (section 2.3.1), hence making it less likely that such DBAs will be afforded the opportunity to attend training from multiple vendors given the cost and time associated with it. Some of the major vendors in the RDBMS environment such as Oracle, IBM, and Microsoft (Moore, 2010), have extensive documentation available online, making it convenient for DBAs to continue their upskilling on areas they are less knowledgeable on, outside of the classroom environment as well as find solutions to problems they encounter daily. Many DBAs also incorporate self-study into their training, making use of online user forums whereby they can find potential solutions from individuals with varying degrees of expertise.

DBAs also consider training courses and associated certifications provided by database vendors more valuable than the benefits of having a tertiary qualification. While respondents acknowledge the broader benefits of having a tertiary qualification, they do not consider it as a mandatory requirement to perform as a DBA. Training received from database vendors are more directly associated with the technical skills they need to perform as a DBA, and hence they ascribe more value to such training courses as opposed to a formal degree. It is thus reasonable to suggest that there is no single method of getting the skills required for the role, but a multifaceted approach is required.

6.5 Theme - Adoption of big data

The *"Adoption of big data"* theme shows the extent to which DBAs in this study have been exposed to big data within their organisations. It also details their current interest in big data as well as their future career aspirations regarding a career in big data. (A summary of the findings for this theme can be found in section 5.5.5).

There is evidence to suggest the given the current lack of exposure to big data that respondents in this study experience within their organisations, it has had a negative influence on their interest in the area and willingness to pursue a future career in big data.

None of the respondents was involved with or was managing big data environments within their organisations (section 5.5.1). This was explicitly expressed by nine respondents who reported having no experience or exposure at present in big data. Only respondent R5 reported that he is managing big data to a degree because of the large volumes of data he must manage. It is then reasonable to infer that there is no requirement at present within such organisations to manage and or invest in big data solutions.

This lack of exposure to big data intrigued the researcher given that the concept of big data started making an appearance as early as the 1990's (section 2.4.1). Big data is becoming increasingly important within organisations and is no longer being discarded but considered as an asset. It presents an opportunity for organisations to extract value from this data and enhance their offerings to their customers given they have the correct technologies and tools to do so (Almeida, 2017). Organisations have thus started making provision for the accommodation of big data within their business by investing in new infrastructure (section 2.4.5). Part of this investment in new infrastructure would be the requirement for database systems capable of handling the requirements of big data because the unstructured nature of big data renders traditional RDBMS unsuitable for this purpose (section 2.4.6). Part of this investment also includes the upskilling of staff with the necessary skills to manage and administer such environments (Luna, Mayan, García, Almerares & Househ, 2014; Ridge, Johnston & O'Donovan, 2015). It is then reasonable to consider that DBAs and other staff would require a level of involvement in getting to understand the new database technology requirements.

When looking at the South African context, one possible reason for the lack of big data exposure reported by respondents can be explained by a study conducted by

Ridge *et al.* (2015). The aim of their study was “to assess the use of big data analytics in the retail industry in South Africa” (*ibid.*:694). Even though their study was not in any way focused on DBAs or their involvement in big data, they reported that the retailers sampled were still primarily focused on the value received from the analytics of structured data within their business and hence not making use of big data as per definition at present. Since a big data technology stack will consist of both relational and non-relational databases, it would require the technical skills of a DBA. Therefore, because organisations are continuing to invest in the RDBMS, it will result in DBAs not having the opportunity to expand their skill set into the big data environment. This could further explain why respondents have not developed an interest in big data since they do not have any big data exposure currently in their workplace.

The reality is that the traditional RDBMS is still the predominant database technology in many organisations, limiting the use of big data database technologies such as NoSQL to larger organisations (King, 2015). Many DBAs are thus still heavily focused in the management of relational databases and given the pressurised environment in which many find themselves in (section 5.2.7), they potentially have little time to invest in researching new database technologies introduced by big data. This is evident through almost half of the respondents in this study reporting a limited knowledge on big data whereby the only knowledge they have is what they have read. This was emphasised by respondent R8 who reported that she was introduced to big data in preparation for the interview while respondent R12 reported that she had only heard of big data. Adriana and Holanda (2016) similarly found there was a distinct lack of knowledge amongst the respondents when it came to NoSQL database technology and big data concepts in general. It is thus possible to understand why almost half of the respondents in this study explicitly showed strong support for the continued existence of relational databases going forward and very few had an interest in big data.

While big data has the potential to unlock new revenue streams, its implementation does not come without challenges which include technical, functional and personnel (Almeida, 2017). The introduction of big data into the business requires functional units from across the organisation to integrate their data and skills, and as such, data management is no longer under the custodianship of a single department (Galbraith, 2012). Business and IT are no longer seen as separate entities but need to combine to convert the data to extract meaningful and valuable information (section 2.8). Galbraith (2012:11) further states that big data has an impact on the organisational

design because various functional units now must integrate their data sources to create a "data network". This introduces a level of disruption from the traditional way in which organisations are accustomed to working. Almeida (2017) supports this view in that regardless of organisational size, implementing a big data solution is complicated because it is not confined to a single department but requires change across the organisation. Thus, many companies are finding it a challenge to incorporate this new potential value stream into their business.

6.5.1 Summary of the Slow adoption of big data theme

The implementation of a big data solution is complexed and requires organisations to not only invest in new technology but also requires a change to their organisational practices, concerning the different functional units across the organisation working together as a single unit. Thus, it requires change across the organisation, and this is potentially too costly for organisations to consider. As such, the business of organisations is still largely being driven by the continued value being derived from structured data and the infrastructure they have invested in to manage such data. Moreover, while NoSQL database technologies have been around for some time, it is only recently that organisations have started adopting it (King, 2017). It is thus reasonable to suggest that the DBA's focus for some time will still be towards the skills required to manage and administer an RDBMS environment.

6.6 Theme - Skills and responsibilities (Big data environment)

The *"Skills and responsibilities (Big data environment)"* theme provides an insight into what respondents consider to be the skills and responsibilities of a DBA in a big data environment. (A summary of the findings for this theme can be found in section 5.6.4).

The discussion of this theme will take place along two threads:

(i) Skills requirement in a big data environment.

(ii) Responsibilities within the big data environment.

The discussion on the skills required by DBAs in a big data environment will be based upon a study conducted by De Mauro, Greco, Grimaldi and Ritala (2017). They identified four job families or roles required within the big data environment through the analyses of more than 2700 job advertisements, grouping various keywords (skills) into skill sets and then associating job families to the different skill sets (see Figure 6-1 below). Based on the findings of this study, this research will

attempt to map the current skills reported by respondents in this study, including the additional skills they consider is necessary to the skill sets as illustrated below.

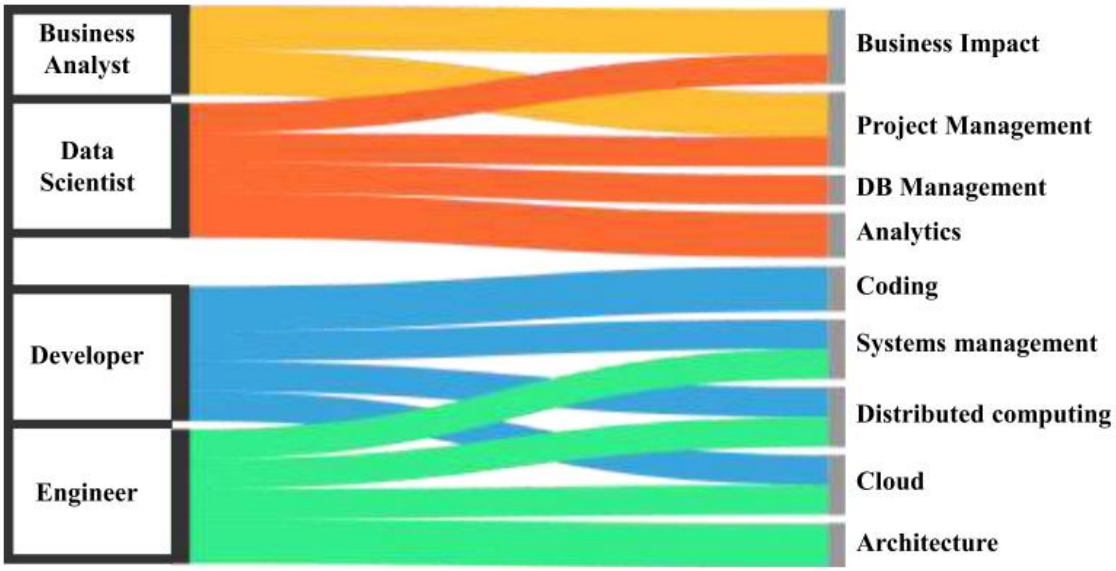


Figure 6-1: Alluvial diagram of Big Data job families vs Big Data skill sets
(De Mauro et al., 2017)

6.6.1 Skills required in the big data environment

The findings suggest that the skills DBAs consider most important to have when transitioning to a big data environment relates to the management of data. While DBAs in their present role manage the organisation's data, respondents in this study reported that DBAs would need to enhance their skill set primarily to include data analytics and data mining.

One of the challenges faced by the RDBMS is that it is not capable of handling the requirements of big data (section 2.4.6) and these challenges present themselves through the characteristics of big data such as volume, velocity, variety, and value (section 2.4.4). Almeida (2017) added another “V” was – “veracity”. Given the diversity of inputs into big data, Almeida (2017:12) refers to this “V” as “the messiness or the trustworthiness of the data”. Madden (2012:4) describes big data as “data that’s too big, too fast, or too hard for existing tools to process”. Not all these big data characteristics pose a challenge for the RDBMS, as great strides have been made to meet the volume and velocity at which data is being generated, and many DBAs are already exposed to managing large datasets. The concern is not with the volume and velocity of data but the variety of data being produced (Abadi *et al.*, 2016). A new set of tools with associated skills is thus needed to accommodate the requirements of big data.

The ability to extract value from big data cannot be managed with a single set of tools but requires a combination of different tools and technologies (Abadi *et al.*, 2016). This makes the big data environment challenging and includes a range of technologies, each fulfilling a different purpose in the creation of a big data solution as illustrated by the NBDRA (see Figure 2-2 in section 2.5). From the NBDRA, a big data technology stack can be viewed along its horizontal and vertical axis. The horizontal axis relates to the different stages of processing the data. These stages include the collection of data from different sources, preparing and formatting data, analytics to convert data to useful information, visualisation as a presentation layer and finally the consumption of information generated (NBD-PWG, 2015). The vertical axis incorporates the different framework components and technologies which support the horizontal processing of data. This includes all the infrastructure resources (e.g. networking, storage, computing, data centres), data storage technologies (e.g. file-systems, relational and non-relational databases such as NoSQL) and processing frameworks for data manipulation using algorithms (e.g. MapReduce) (NBD-PWG, 2015). While this is a simplistic overview of the big data environment, it still highlights the complexity of the environment needing a diversity of skill sets.

In order to manage this new complex environment, the role of the data scientist has been defined as a role with diverse technical and business skills, and having many responsibilities within the big data environment (section 2.9), i.e. someone who is multidisciplinary and considered an all-rounder. While many considered this role to be an all-encompassing requirement to manage big data, De Mauro *et al.* (2017) state that differentiation should be made between the field of data science, and the role of the data scientist relating to skills required in a big data environment. They reported that all the requirements of data science could not be fulfilled by a single role, but several different skill sets are required to transform big data into information and thereafter derive business value. They defined four job families as such within the big data environment namely, business analysts, data scientists, big data developers and big data engineers (De Mauro *et al.*, 2017). Thus, not all the skills needed to derive value from big data are resident within a single role. This is further illustrated in that for organisations to derive business value, various departments with various skills must integrate, creating a data network within the organisation through the field of data science (section 2.8).

The primary skills reported by respondents in this study related to the requirement for DBAs to improve their skills needed to convert data into useful information such

as data analytics and data mining. While data analytics and data mining are key to extracting value from big data (section 2.8), these are not competencies that generally fall within the skill set of DBAs, but are the core competencies to be fulfilled by the data scientist (section 2.9) (Davenport *et al.*, 2012; Zhu, Xiao, Zhang, Gu, Yi & Li, 2015). This is also illustrated and supported by De Mauro *et al.* (2017) who report that the data scientist's core skill relates to analytics while other competencies include understanding the business, project management, and database management skills. Working from the keywords underlying the skill sets and levels of proficiency De Mauro *et al.* (2017) used to map job families to skill sets, the findings of this research show that DBAs have a strong match with the level of database management skills required by data scientists but they do not have skills in the core competency of the data scientist, i.e. analytics. As a result, should they consider upskilling in analytics, it will provide them entry towards the role of the data scientist although they will not be sufficiently skilled to fulfil the role in its entirety. Data analytic and data mining skills would include tools such as Apache Mahout for machine learning and data mining, while Apache Hive is used to query and manage datasets (section 2.6.5).

Respondents view on the need to upskill in cloud technologies is consistent with the view of Abadi *et al.* (2016:94) who stated that "many big data applications will be deployed in the cloud, both public and private, on a massive scale" and there is currently a definite move towards database systems being hosted within the cloud (King, 2017). Cloud computing has also shown many advantages over on-premise database systems. These advantages include savings regarding the cost of the service through the pay-as-you-go model, reduced deployment time while taking care of other complex functions like scalability and high availability (section 2.7) (Holt *et al.*, 2015). Increased attention has since been given to the cloud for its infrastructure, scalability, flexibility and payment model to accommodate the distributed processing requirements of big data (section 2.7). With reference to the job families defined by De Mauro *et al.* (2017), skills relating to cloud technologies are primarily within the roles of the big data developer and big data engineer. Many DBAs have though, already started gaining exposure to cloud technologies through services such as DBaaS. It has thus shifted their attention to managing databases both within the cloud as well as those still resident on-premise (section 2.7). Therefore, cloud technologies will be a skill that DBAs will need to incorporate into their skill set even if they do not transition towards a role within the big data environment.

Many respondents in this study have prior experience in development roles and have reported programming as a key skill required of DBAs. A large part of their role relates to interacting with the database through SQL, as well as supporting software development teams, and having knowledge of development languages also aids them in this regard ('Boost your career as a database administrator', 2013). In recent years, the introduction of Development and Operations (DevOps) within many organisations has also resulted in DBAs being more involved in the development process (King, 2017). The requirement for programming skills can be further explained through many NoSQL and NewSQL database vendors incorporating their own query language to access the data (section 2.6.1, 2.6.3). This adds a level of complexity, given that there is no standard yet across the different vendors. However, many vendors have started to incorporate SQL as a query language, and this would, in turn, focus on one of the DBA's key skills. Coding is identified by De Mauro *et al.* (2017) as an important skill set of the big data developer, and here again, there is an overlap in terms of the skills within the traditional DBA's current skill set, as well as what respondents in the study considered to be an important skill to have in a big data environment. Even though coding is considered a core skill of the big data developer, the keyword "programming" appears across several skill sets. However, the level of proficiency required within the other skill sets is much lower than the proficiency required by the big data developer. It is, therefore, considered an important skill, and it puts the DBA in a good position in relation to this skill.

Mapping DBA skills to the skill sets identified by De Mauro *et al.* (2017)

Figure 6-2 depicts a mapping of existing and additional skills reported by respondents in this study to the skill sets as reported by De Mauro *et al.* (2017).

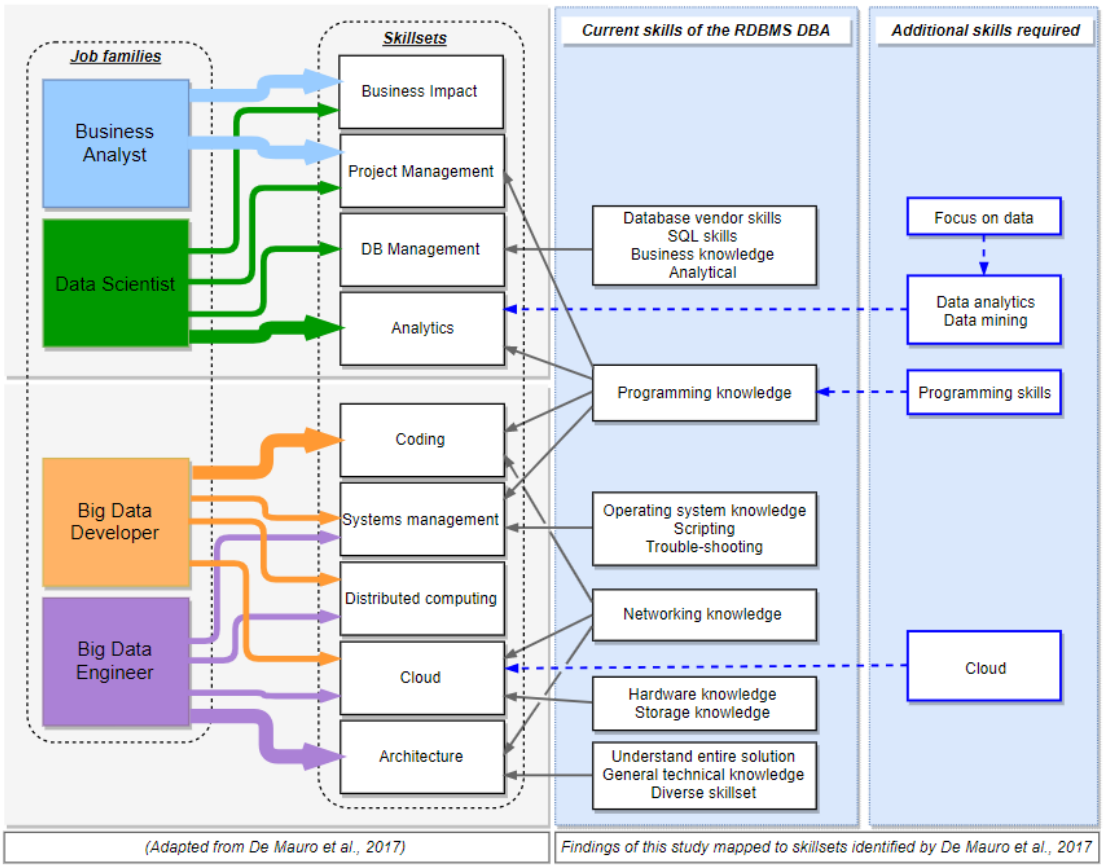


Figure 6-2: Mapping the findings to De Mauro *et al.* (2017)

De Mauro *et al.* (2017) used a number of keywords to create their skill sets, and these same keywords have been used to map skills identified in this study to the skill sets of *their* study. These mappings were based on the researcher's interpretation to create the best possible match. As can be seen in Figure 6-2, many of the DBA's existing skills map to several skill sets and subsequently indirectly link to different job families. Each of the skill sets of their study consists of many different keywords (skills), and the association of a skill in this study to a skill set in their study does not mean that the DBA meets all of the requirements of that skill set. In many instances, it only provides entry to that skill set. While this is not perfect, it shows that the DBA meets many of the various skill set requirements, but not enough to meet them fully. This then filters through to the job families, whereby the DBA partially meets the skill set requirements of three of the four job families namely the data scientist, big data developer, and big data engineer. So, because of this widespread mapping across the skill sets, there is no one particular job family in which the DBA role fits.

The following is a brief description of Figure 6-2 to show how the DBA's skills map to the skill set requirements of the four job families:

Business analyst:

(i) Weak mapping though programming skills to the project management skill set.

(ii) No mapping to the business impact skill set.

Data scientist:

(i) The DBA's database management skills fully meet the database management skill set requirements.

(ii) No mapping to the core competency, namely analytics.

(iii) No mapping to business impact and project management.

Big data developer:

(i) Maps to three of the four skill sets namely, coding, systems management, and cloud but only partially meets the requirements of each of these skill sets.

(ii) No mapping to distributed computing which is an important part of big data.

Big data engineer:

(i) Maps to three of the four skill sets namely, systems management, cloud, and architecture but only partially meets the requirements of each skill set.

(ii) No mapping to distributed computing which is an important part of big data.

6.6.2 Responsibilities within the big data environment

Findings of this study reported that almost half of the respondents are unclear as to the role of the DBA within the big data environment while to a lesser extent some reported it to be a similar role. There is no clear evidence as to what the responsibilities of the DBA would be in a big data environment. The responsibilities will also be dependent on the environment in which the DBA finds themselves in, such as on-premise versus the cloud.

Big data allows organisations to shift away from the analysis of static and historical data, enabling decisions to be based on a continuous flow of data in real-time (Davenport *et al.*, 2012). This has imposed a potential new challenge for DBAs. For organisations to exercise this new ability and for their decisions to be effective, DBAs now more than ever must ensure that such data is available and responsive to provide that data whenever required.

It is the researcher's view that the respondents reporting the need for less technical skills can be strongly associated to the automated and self-management features

being introduced by database vendors, as well as the adoption of the cloud within organisations. Given the size and complexity of a big data solution, the infrastructure would also need to incorporate self-management features whereby hardware failures and other outages will be taken care of without the need for human intervention (NBD-PWG, 2015). As such, this will also have an impact on the responsibilities to be performed by DBAs, whether they continue to manage the RDBMS environment or transition towards a role in the big data environment. Cloud computing has provided flexibility and automation of many previously complex and time-consuming functions which in turn require technical expertise (Holt *et al.*, 2015). The requirement for DBAs to also incorporate knowledge of supporting infrastructure into their skill set and hence a level of responsibility in this regard will also be minimised. With cloud being a common supporting platform for big data solutions, many of the responsibilities performed by the DBA would now be taken care of by the cloud vendor, depending on the service provisioned such as IaaS, PaaS or SaaS (section 2.7). This is emphasised by Abadi *et al.* (2016) who states that all the administrative tasks in a PaaS environment are automated. Hence there is less need for DBAs. However, its low cost and scalability make it an ideal platform for the processing of big data (Almeida, 2017) and the adoption of cloud services is on the increase within organisations (King, 2017).

There is no clear evidence as to what the DBA's responsibilities are within the big data environment. As stated in the literature (section 2.10), there is now an overlap with other roles within the big data environment such as system administrators, network administrators, developers, and analysts. This is further illustrated by De Mauro *et al.* (2017), who defined four job families within the big data environment, none of which explicitly show the role of the DBA (see Figure 6-2). As such, it is difficult to clearly understand what the functions and responsibilities of the DBA role will be. While King (2017) reported that DBAs would be responsible for managing new database technologies such as NoSQL, the same study also found that these databases could often be managed by other teams such as application teams. It is then possible that the DBA's primary responsibility could still be managing the RDBMS environment within the organisation. Many organisations are still heavily focused on the value derived from structured data within their business. Structured data within organisations is still on the increase and with the added requirement for DBAs to manage databases from different database vendors (King, 2017). This was to a degree supported by Almeida (2017) who reported that many organisations are still finding it challenging incorporating big data into their business.

6.6.3 Summary of the Skills and responsibilities (Big data environment) theme

The big data landscape is complexed and requires a diverse range of infrastructures, technologies, and processes to manage, process, and derive meaningful insights in order for organisations to add new revenue streams to their business. As such, it requires a diversity of skills to accomplish this, not from a single individual but incorporates various skill sets from across the organisation. In relation to the job families and associated skill sets identified by De Mauro *et al.* (2017), there is no clear evidence to suggest where the DBA skills will be required within the big data environment. Many of the current skills that DBAs have overlap with different skill sets across different job families. In addition, the primary skills reported by DBAs in this study regarding data analytics, data mining and focusing on data are identified with the important competencies of the data scientist. Many of the secondary skills required by DBAs such as knowledge of infrastructure supporting the database, will now also be taken care of by the service provided by the cloud. As such, it is also not clear what the role of the DBA would be, and it is possible that the DBA self could potentially fulfil an all-rounder type role or continue in the management of the RDBMS environment.

6.7 Theme - Big data career pathing

The *"Big data career pathing"* theme shows that DBAs in this study has very little interest in pursuing a career in big data and that the large majority of respondents are seeking careers in other fields within the IT sector (A summary of the findings for this theme can be found in section 5.7.4).

There is strong support for DBAs not wanting to pursue a career in big data but to continue towards other careers within the IT sector. Potential contributing factors can be associated with the lack of opportunities to explore big data within their organisation through the continued organisational interest in extracting business value from structured data. The DBA role can, therefore, be considered as not being a life-long career but an intermediate stepping stone towards other careers.

Since the DBA role is technically inclined, and given that the complexity of a big data environment will provide DBAs with an opportunity to extend their technical ability and interest, one would expect that this would be a career that DBAs would want to pursue. DBAs are inspired by their ability to experience new technology, and it is possible that should the respondents in this study have experienced such technologies in the workplace that they might have had a different perspective about big data. DBAs wanting to pursue other careers is not unique to this study though.

While this study is not assuming that the databases being supported by respondents in this study are outdated, a study by McKendrick (2014) reported that many professionals are seeking opportunities to expand their technical skills on cutting-edge technologies and they are less inclined to want to maintain older systems. Big data technologies would thus provide the opportunity for DBAs to experience cutting-edge technology, but it is evident from the findings that there are currently no opportunities for DBAs to explore big data within their working environment. This was reported by respondents by either indicating that they are not aware of any career development plans towards big data or that they are unaware of the organisation's intention to move towards big data.

The desire to pursue a career away from the DBA role can further be explained by the current working environment in which DBAs work. While a pressurised working environment is not unique to the DBA field, respondents still described the DBA environment as highly pressurised. Added pressures such as potential financial implications should SLAs be breached and having to troubleshoot issues that arise in limited and stressful times can become contributing factors to seeking solace in other professions. As a result, even though DBAs do not currently have opportunities to explore big data within their organisations, they potentially also have little opportunity to explore other avenues of interest such as researching new and emerging technologies. This was again highlighted by McKendrick (2014) who reported that many DBAs were becoming increasingly frustrated with the lack of opportunities to learn new skills, the reasons being the increased complexity of database systems and a large proportion of their time is spent managing various crises.

People make different choices about their career development. While some individuals have a plan and pursue it carefully, others will seize opportunities presented to them (Matthews, 2017). The findings of this study have also shown that respondents have traversed various roles within IT towards the DBA role, either intentionally, opportunistically or through no choice of their own (section 5.3.3). That respondents desire to further transition towards other roles within IT shows that it is possible to consider the DBA profession not as a life-long career, but as a means to an end towards other professions within IT. Given that individuals have the skill set to take up other opportunities, many individuals take risks and explore where it leads them (Matthews, 2017).

6.7.1 Summary of the Big Data career pathing theme

There is apparently a decline in the number of DBAs going forward as many are exiting the profession in the near future, while others are looking to move towards a career in data science (McKendrick, 2014). Many reasons can be attributed to this decline or lack of desire to continue in the field of database administration. As this study reporteds, there are many factors that can lead to this exodus, such as the desire to learn new technical skills, the lack of opportunities presented through the continued use of the traditional RDBMS, and continuous pressure experienced in the DBA environment. DBAs are therefore becoming frustrated by not being able to expand their skill set by experiencing new technologies. As a result, they are potentially losing interest in continuing their career within the DBA role.

6.8 Theme - Transitioning to a big data role

The *"Transitioning to a big data role"* theme shows that DBAs are confident in their ability to transition into a similar role within big data. While respondents consider that there will be a requirement to re-skill, it will be an upskilling and enhancing of skills rather than discarding any of the existing skills they have previously acquired. (A summary of the findings for this theme can be found in section 5.8.4).

DBAs are technically skilled professionals who have transitioned through various roles within IT towards the DBA role. They have a diverse skill set, acquired through experience in prior roles and while performing as a DBA. These existing skills will not be discarded but will enhance DBAs skills to manage the tools within the big data environment. They are confident that they can make the transition from the role within the RDBMS environment to a similar role within big data.

It is understandable that DBAs are confident in their ability to transition to a role within big data, given that they are technically skilled individuals (section 6.2). The majority of respondents in this study have made a transition from other roles within IT (section 6.3), bringing along experience, expertise and a range of skills, predominantly from a software development background. These skills will continue to be of benefit to them within the big data environment as it is currently being used within the RDBMS environment.

Almost half of the respondents report a requirement to re-skill in order to transition to a role within a big data environment. The term "reskill" is defined as *"*Teach (a person, especially an unemployed person) new skills*",* whereby the definition of "upskill" refers to "Teach (an employee) additional skills*"* (Oxford Dictionaries |

English, 2018). It is the researcher's view that DBAs will need to upskill rather than re-skill. The current role of the DBA in the RDBMS environment requires an individual with a diversity of skills (section 6.2), and the requirement of a big data environment will be no different. The majority of respondents in this study have also shown that DBAs are educated individuals, supporting McKendrick (2014). It will not be that existing RDBMS skills and knowledge will be discarded, but rather that a new set of skills and knowledge will be needed. The added skills developed for big data will be complementary to the array of skills and knowledge DBAs already have.

Respondents considered vocational training as a means to upskill them towards a role within big data. Respondents' attitudes towards getting certified in big data was consistent with their positive attitude towards vocational training acquired in the RDBMS environment (section 6.4). Many DBAs consider such certifications as an important component of the education and training of a DBA, in addition to it having a positive influence on future employment opportunities (McKendrick, 2014). Because such certification programs are tough and require a level of commitment to complete successfully, it does enhance the DBAs ability to find better job prospects ('Boost your career as a database administrator', 2013). Therefore, like the vocational training and associated certifications received from database vendors such as Oracle, Microsoft, and IBM, DBAs can upskill themselves in big data. Learning paths and certification exams are currently available for NoSQL products such as Cassandra and MongoDB (Simplilearn.com, 2018; University.mongodb.com, 2018). In the light of DBAs favouring the technical guidance received from such training, it is evident that this will be an avenue explored in conjunction with the current methods they use to upskill themselves (section 2.3.4, 5.4.4).

Financial pressures limit the ability of many organisations to provide extensive training opportunities to DBAs, and they are thus looking for individuals who are already suitably equipped with the necessary skills (Simmonds, 2013). The complexity of the technology stack in a big data environment could further complicate the upskilling of DBAs as it could potentially require upskilling for a number of technologies. As such, organisations face a complex task to first decide upon which big data solution will support their requirements and as well as the individuals with the appropriate skills to manage such solutions (section 2.5).

6.8.1 Summary of the Transitioning to a big data role theme

The findings of this study suggest that DBAs are largely confident in their ability to transition to a role within the big data environment. They are educated and

technically skilled individuals, who have been able to transition from other roles within IT to take up a position as a DBA. The skills they have thus acquired in previous roles, as well as those acquired while performing as a DBA, stands them in good stead to continue their upskilling towards a role within big data. Re-skilling in big data will not be a requirement as existing skills and knowledge will still be applied within the big data environment, but an upskilling towards big data will be required to further enhance the skill set.

6.9 Summary of Chapter Six

Chapter Six provided a detailed discussion of the themes that emerged from the findings presented in Chapter Five in relation to existing literature.

DBAs have shown themselves to be technically skilled and educated individuals. The diverse skill set includes both technical and non-technical competencies needed to perform the many duties required. This skill set can not only be attributed to the requirements of the DBA role but also to the extent that many DBAs have transitioned from other roles within IT, mainly having a background in software development. It is a role that requires extensive collaboration, in both providing and receiving services from other departments within the organisation. The importance of the DBA role is further highlighted through the importance many organisations place on the integrity and availability of their data to conduct their business.

There is no single method DBAs use to upskill themselves. While training received from database vendors received positive support to a large extent, it is not always possible in terms of cost and time. DBAs often support multiple vendors and must troubleshoot issues regularly. They must, therefore, be self-sufficient when it relates to their knowledge and upskilling, making use of a wide variety of methods such as the internet, workshops, conferences, vendor training, and online documentation to name a but a few.

The adoption of big data is not as widespread with none of the respondents currently working with big data in their organisations. Their primary focus is still around managing an RDBMS environment given that many organisations still find value in the structured data within their organisation. Hence, interest in big data was low with the majority foreseeing themselves in other careers within IT in the near future. A contributing factor to this exodus can also be associated with the pressurised working environment in which many DBAs find themselves in. Despite this lack of

exposure to big data, DBAs were nonetheless confident in their ability to make the transition should it be required.

The big data environment is complexed and requires a diversity of skills in different roles, each fulfilling a specific function. Respondents reported the need to upskill in data analytics, data mining, cloud, and programming. However, many of these additional skills they reported along with their current skills overlap with many roles identified within the big data environment. As such, there is no clear evidence to suggest where the DBA role fits into this environment. This also brings about some uncertainty regarding the role and responsibilities of the DBA, whether they will fulfil an all-rounder type of role or continue managing the RDBMS environment.

The conclusions in relation to answering the research questions to be addressed by this study will be presented in Chapter Seven. The chapter will also include the recommendations made by the researcher.

CHAPTER SEVEN: CONCLUSION AND RECOMMENDATIONS

7.1 Introduction

Chapter Six discussed the seven themes that emerged from the findings presented in Chapter Five. The aim was to determine the current skill set of DBAs working in an RDBMS environment and the skills required to enable them to transition to a similar role within a big data environment. The primary objective of this chapter is to provide answers to the research sub-questions in support of answering the main research question stated in Chapter One.

The research problem this study is trying to address is the shortage of specialists to administer big data environments in organisations. Even though DBAs manage the organisation's traditional RDBMS, they may not have the required skills to work in a big data environment. Organisations might therefore not be able to extract the potential value that big data holds if they do not have the necessary skilled staff to manage and analyse big data within the organisation.

The conclusions for this study are discussed in relation to the research questions to be answered and to determine to which extent the objectives of the study have been achieved. Based on the findings and subsequent discussions of said findings, recommendations are presented along with possible areas for future research. The limitations of this study are also considered and described.

7.2 Research aim and objectives of the study

Research aim

- To explore the difference in skill set between a DBA administering an RDBMS and a similar position in a big data environment.

Research objectives

- To explore what technical re-skilling will be required of DBAs when transitioning from a position within an RDBMS environment to that of a big data environment.
- To develop a model to assist organisations in accommodating the DBA role within the big data environment.

7.3 Mapping of skills to skill sets

A reworked model of that presented in Figure 6-2 can be seen below in Figure 7-1. The findings of this study have been mapped to the skill sets as presented by De Mauro *et al.* (2017). The intention of the illustration is to identify which job families the DBA can potentially transition towards, and this is done by mapping the skills DBAs consider they need and their current skills to the skill sets associated to the different job families.

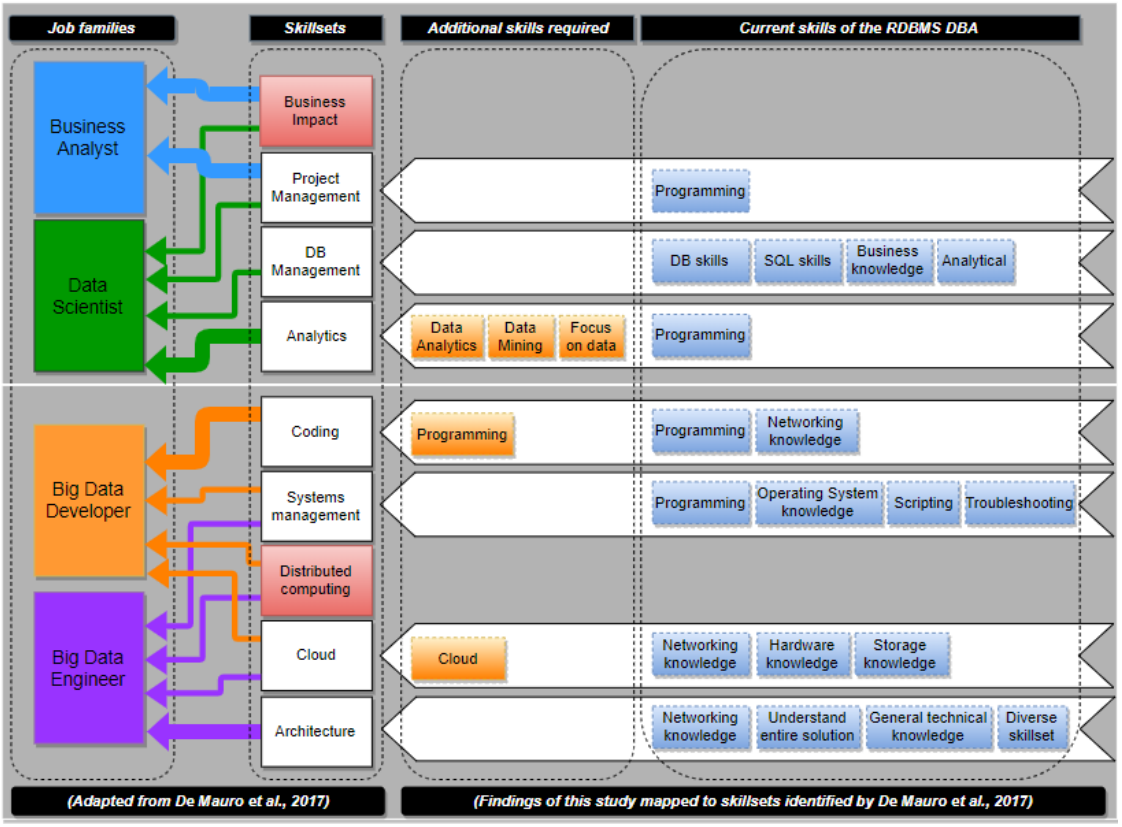


Figure 7-1: Final model of skills requirement for DBAs in a big data environment

7.4 Answering the research sub-questions

Research sub-question 1: What are the current skill set and responsibilities of a DBA working in an RDBMS environment?

Skill set of the DBA

There is no standard template which can be used to describe the skills and responsibilities of DBAs. While there are core competencies, skills, and responsibilities required to administer the RDBMS, it is evident that the DBA's skill set is broader than just having knowledge related to the database software for the different vendors they support. The findings of this study have shown that the DBA role requires an individual who has a wide range of both technical and non-technical

skills while the scope of the DBA's responsibilities has broadened to be more than just the management of databases.

The DBA's core competency is the administration and management of the database for the different vendor's solutions being supported, but they also require knowledge and understanding of the infrastructure upon which the database is built. The added complexity of supporting multiple vendors requires knowledge of different operating systems, for example, Linux, Unix and Windows as well as networking and storage. Knowledge and expertise of these components are not because DBAs are required to manage and administer them, but it aids the DBA in troubleshooting and being able to identify issues that arise. Performance related issues are not always related to the database self but can often include external influences on the database such as inefficient SQL, poor disk I/O potentially related to hardware or storage, and or network latency. It is thus important for the DBA to have some understanding of these components to enable them to communicate with and work alongside the specialists within these areas to resolve such issues.

The skill set of many of the DBAs in this study has been enhanced and influenced through their experience in other roles before becoming a DBA. Often, individuals progress *towards* the role of a DBA rather than *pursuing* it as a first-choice career. Many have transitioned from other roles within IT, and in some instances, they have transitioned across multiple roles. Common amongst many of the DBAs in this study was exposure to software development, either having transitioned directly from a software development role or having occupied a role as a software developer somewhere in their career. This helps them in many respects because they often must help and mentor software developers by troubleshooting SQL and other application related errors. This also extends the DBA's skills not only to be focused on the database software but also to broaden their skills to incorporate knowledge and understanding of the entire solution.

Through their transition from other roles, many DBAs bring with them tertiary qualifications in either CS or an IT stream. There is no compulsory requirement to have a tertiary qualification to perform as a DBA. These qualifications also do not have a direct influence or contribution to their technical skills to manage and administer a database but contribute to their broader knowledge and expertise as an IT professional. It is the foundation upon which their diverse skill set and broad technical knowledge is built. Therefore, because of experience and qualifications

obtained through other roles, many DBAs bring with them skills and expertise which aid their ability to perform as DBAs.

A diversity of sources also influences the skill set of the DBA. They are self-taught in many respects as a large proportion of their upskilling happens through self-study. Where financially possible and time allows, the speciality technical expertise received from database vendor training contributes positively to their technical skills and knowledge of the database software. The skills are primarily focused on the effective management of the database software. As such, training received from database vendors is largely favoured amongst DBAs, and it is considered to having a positive influence on their skill set.

While being technically skilled in managing and administering the database solution, it is imperative to recognise the importance of non-technical skills. DBAs do not work in isolation, but it is a role that is visible across the organisation, working as part of a team of different functional units to support the business. They support many of the organisation's functions, both directly and indirectly. Therefore, their variety of skills contribute to their ability to work with different stakeholders within the organisation, ranging from software development teams to senior management. These skills include communication, teamwork, various social skills including an advisory and mentoring role to software development teams with whom they have a close working relationship. So, because the DBA environment can often be demanding, the ability to manage different pressures from a technical and social perspective is indispensable.

Responsibilities of a DBA

The DBA role is a specialist role within IT and requires skilled individuals to undertake the responsibility to manage and administer the organisation's data. Whether they administer development or production environments, the data within those databases is a valuable commodity, enabling the organisation to derive value and drive their business. The role of the DBA should thus not be underestimated since many organisations have a financial dependency on the availability and integrity of their data. Such data is only of value to the organisation if they can access it efficiently and reliably.

The primary concern of the DBA is to ensure the effective administration of the database systems for which they are responsible. Minimising any potential downtime is achieved through various means and includes continuous monitoring, ensuring

that valid backup and recovery solutions are in place and practised, and to ensure that availability is maximised. Many of these tasks involve both the database as well as the underlying infrastructure which supports the database. While DBAs are not responsible for resolving issues that arise from underlying infrastructure components, availability of the database is still dependent on these components. Availability thus also incorporates the ability of the database to service end-user transactions as efficiently as possible. Even minor degradation in network performance or disk I/O can negatively influence the performance of the database, and subsequently, have a financial impact on the organisation's business in certain environments. The ability to automate the monitoring functions to ensure the health of the database system is thus important, because such tasks are mundane and time-consuming, and can potentially be neglected if performed manually.

While it is important to provide access to data, it is critically important to safeguard the data to ensure that only authorised users have access. The DBA must also ensure that functions such as security, auditing, regulatory compliance, and governance form an integral part of the DBA's environment. Sensitive data such as organisation financials or personal user information is often only accessed by specific individuals within the organisation. This requires DBAs to adopt the principle of least privilege to ensure that the integrity of the data is secured, and that data breaches or leaks are prevented. The DBA self must also adhere to these security requirements as they are the data custodians and should thus not abuse their authority to gain access to information they are not allowed view.

The need to minimise disruption to the organisation's business often results in a large proportion, if not all the DBA's maintenance activities being performed after-hours. Software installations, upgrades, patching, and configuration of databases must be performed when it least affects the end-users. Many of these environments run 24 hours a day, 365 days a year such as online retailers, making it even more crucial that these activities be streamlined as much as possible. Downtime is often scheduled for a fixed period, and these activities must be completed within that period. As a result, it is important that DBAs practice these activities in non-pressurised test environments, to ensure they remove all potential issues that could arise and ensure the process is completed timeously.

The scope of the DBA's responsibilities is broader than just the administration of databases. They are central to many functional units and stakeholders within the organisation. These range from the strong working relationship with software

developers, to interactions with senior management. DBAs often assist software developers with troubleshooting problematic or complex SQL statements, providing guidance and mentoring. There are also instances whereby DBAs must further extend their job description with responsibilities not associated with the role of the DBA. It is thus a role that is diverse in both skill set and responsibilities, and it is often that these two are complimentary.

Research sub-question 2: What are the skills and technology requirements of a big data environment?

Skills requirements

A big data environment cannot be managed by a single role or skill set but needs a range of skills across the organisation to derive business value through the field of data science. Data is generated from disparate sources, from outside the organisation but internally as well. Many functional units or departments within the organisation must now merge their data sources to create a data network in conjunction with data generated externally of the organisation. Business and IT no longer function as separate units but must integrate and as such, several functional units within the organisation must merge their data and skill sets to unlock the potential value that big data holds.

While many consider the role of the data scientist as a role that could meet the skills requirements of big data, it has been reported that the data scientist fulfils only a part of the big data skill set requirements. Their primary function is data analytics, using different techniques and algorithms to find patterns within data and develop business models to enable the organisation to unlock potential new revenue streams. The field of data analytics, however, is broad and includes different functions such as data mining, machine learning, predictive analytics, statistics and more. While respondents in this study considered the need to upskill in data analytics, data mining, in addition to programming and the infrastructure supporting the big data environment namely, cloud computing, it is much more than that.

Processing big data needs new skills to manage and administer the distributed processing frameworks such as Hadoop which incorporates MapReduce and HDFS. This also now includes YARN, a cluster resource management framework, which is a new feature within Hadoop 2.0. Hadoop's open-source architecture is the basis of many custom implementations which different solutions providers have tailored to meet their requirements. Skills to manage and understand the distributed

processing, computing, and resource management of data across potentially thousands of commodity servers is a requirement and includes a strong need for programming skills.

Data storage skills in relation to non-relational database technologies such as NoSQL and file system storage is needed. While relational databases incorporate functionality such as joins and constraints, amongst others, to maintain the integrity of the data, non-relational databases do not. The onus is now placed upon the application to incorporate these functions as part of the code. The requirement to understand programming is further highlighted by different solution providers including their own custom query languages to access the data. NoSQL databases themselves have several variations such as key-value store, column store, document store, and graph databases, each meeting a specific requirement. There is also a need for continued skills for managing relational databases because non-relational databases are not a replacement. SQL skills will therefore still be a requirement as many NoSQL database technologies have started incorporating this as a query language, while tools such as Hive provide a SQL-like language, HiveQL, to manage large datasets.

While all these different technologies form an integral part of the big data technology stack, the architecture, and infrastructure that supports this big data environment needs to be managed. A platform that can meet the scalability and flexibility requirements of big data is often met by the infrastructure and services provided by cloud computing. While there is a requirement to manage the cloud environment, the skills and level of involvement in managing these environments are dependent on the type of service provisioned. These services include IaaS whereby the cloud vendor manages the infrastructure only, PaaS whereby the cloud vendor manages everything except the applications and SaaS whereby the cloud vendor is in complete control. Many of these services and functions are automated within the cloud vendors offering and thus lessons the technical skills requirements and responsibilities of DBAs and other technical roles within this environment.

Technology requirements

Big data needs an integrated technology stack, combining different technologies to create a solution. Considering the NBDRA (see Figure 2-2), there are several layers upon which a solution is built. The first layer is the supporting infrastructure, and this is often met through the services provided by cloud vendors. These are the physical resources within data centres, consisting of a range of commodity hardware

configured within a cluster. Cloud-based environments provide scalability, flexibility and potentially a cost-effective infrastructure required by big data by the virtualisation of resources. It allows the provisioning of resources on-demand, distributed across several data centres. The infrastructure and services provided by cloud vendors allow the organisation the flexibility of not having the burden of provisioning their own hardware on-premise and thus enables them to concentrate on deriving value from their data.

Data storage is accommodated through database technologies such as relational and non-relational databases. While non-relational databases such as NoSQL and NewSQL have been born to accommodate the need for unstructured data, relational databases are still needed. Non-relational databases are not a replacement for relational databases, but they are complimentary. Relational databases are based upon the relational model, but non-relational databases have a variety of models to accommodate different needs. These include key-value store, column store, document store, and graph databases. Many big data solution providers have since implemented their own variations of these databases to complement their solution, and there is thus no standard implementation. This is also seen through the query language and interface used to access the data. Each vendor has their own programming language which they use to query the database, unlike relational databases whereby ANSI SQL is a common implementation. Many vendors have since started incorporating SQL as a query language to achieve some standardisation across the different technologies.

Distributed processing or computing is met through the specific frameworks to accommodate the enormous volumes of data to be processed within big data. The ability to scale on-demand using commodity hardware is of fundamental value. The vertical technology stack of relational environments does not lend itself to scalability and thus makes it difficult. It is often expensive, requiring the replacement of an entire technology stack with a finite lifespan to meet the requirements for which it was provisioned. Horizontal scaling and processing are achieved by processing chunks of data in parallel across several servers, each providing its own processing and storage resources. This allows for adding and removing of resources depending on the need at the time of processing. Common implementations of such distributed processing frameworks are Hadoop which consists of HDFS and MapReduce. These are open-source frameworks which many big data solution vendors have customised to meet their requirements. Hadoop is not the only such framework but is the basis upon which many new variations have been developed. Hadoop 2.0 has since been

developed and incorporates YARN which is a cluster resource management framework. YARN also provides several enhancements over the existing MapReduce framework, which includes improved scalability, parallelism, and resource management.

While there is a requirement to store and process big data, it also requires supporting tools used to create meaningful value from the information derived after such data is processed. These include tools for data analytics and data mining, such as Mahout, using algorithms and models to convert data into information. Other tools such as Hive, used for data management and data warehousing, support the data analytic functions by providing data summarization and ad hoc querying capabilities. The consumption of this data is further supported by the visualisation tools which enables meaningful insights into such information and thereby driving business value.

Research sub-question 3: What career plans do organisations have for DBAs with regards to big data management?

No career plans have been identified within organisations to upskill DBAs towards a role in a big data environment. This is because organisations either have no immediate plans to incorporate big data into their business or they do not know where the DBA role fits within the big data environment. Most organisations are still finding value in structured data, thus slowing the adoption of big data. This is also amplified by the complex nature of a big data environment. It requires a combination of technologies, tools, and skilled staff and brings about change to the organisational structure. This organisational change is brought about through various functional units within the organisation now having to integrate, having an impact on the current business processes and practices.

The lack of career paths towards big data has potentially lead to many DBAs showing a distinct lack of interest in big data while some are considering exiting the DBA profession in pursuit of other roles within IT. Organisations are also finding it hard to define a role for the DBA in a big data environment when considering the four job families defined in the literature. While DBAs have presented a diverse skill set, there is no clear path that they can follow towards any one of these job families because they do not fully meet the requirements of the different skill sets associated to these job families.

7.5 Answering the main research question

Research question: What is the skills gap between the DBA working in an RDBMS environment and a similar role within a big data environment?

The difference in skills between the DBA working in an RDBMS environment and a similar role in a big data environment must be considered along to four job families present within the big data environment (see Figure 7-1). DBAs have a diverse skill set and fulfil many responsibilities within the organisation. Nevertheless, they are not entirely ready to become part of any one of the four job families. Although the skills DBAs presented maps to several skill sets associated with the job families, it is entry level rather than having the complete skill set to enable a transition to one of them.

While the DBA's competency in database administration would meet the database management skill set requirements of the data scientist, they do not meet the core competency needed of the data scientist, specifically analytics. They also fall short of the other skill sets within the role of the data scientist, namely project management and business impact. The latter two skill sets are the core competencies of the business analyst, and the DBA's current skills may not enable them to perform to this role, because none of their current skills would benefit them in this regard. There is thus a stronger association towards the technical job families namely, big data developer and big data engineer and while there is an association towards these job families, DBAs are still sitting on the cusp of the skill set requirements and are therefore not sufficiently skilled at present to transition towards any one of them.

While DBAs have shown that they are capable of transitioning to new roles, it must be emphasised that it is not just a case of re-skilling or adding more skills, but careful consideration must be given as to which job family a progression is made towards. DBAs are well positioned across the big data developer and big data engineer job families. Each of these roles has a specific function and core competency. To enable DBAs to transition to either of these job families they will need further upskilling in several skill sets. Common to both these technical job families are systems management, cloud, and distributed computing. While DBAs have presented some skills within the systems management and cloud skill sets, there is a distinct lack of skills in the distributed computing skill set. This is a core part of any big data environment and is thus a requirement.

While there is a need to upskill, the DBA's current skill set will not be discarded as there is evidence that their skill set will allow them to build upon the skills already required within the RDBMS environment. Furthermore, relational databases will not be discarded because a combination of both relational and non-relational database technologies will be needed. Therefore, should the DBA upskill towards one of these job families, they are nicely positioned to continue managing the organisation's relational database requirements. Hence, they can fulfil more than just the big data developer or big data engineer role.

7.6 Contribution and recommendations

This study has shown that the DBA has the potential to transition towards a role within big data, although this new role would not necessarily be the same as present within the RDBMS environment. Big data has some very specific requirements due to the different components of its technology stack and has shown that it is a more complexed environment than the RDBMS environment. It requires several different roles each fulfilling a specific function. It ranges from the ability to bridge the gap and communicate with organisational stakeholders to derive value, to the roles supporting the technical requirements of big data.

While the DBAs' diverse skill set has positioned them nicely to move towards a role within big data, it is evident that there is a definite requirement for them to upskill and enhance their current skill set. Big data requires skill sets that are specialised in relation to the four job families while the DBA's skill set could be considered as being too broad to meet the skill set requirements of any one specific job family. It is not only a case of adding more skills, but careful consideration should be given to the direction to be pursued within a big data environment. While many of these job families have skill sets that overlap, they each have specific core competencies in which they must have a higher proficiency.

While this research has contributed towards enabling the DBA to move to a role within big data, it is still dependent upon the adoption of big data to provide DBAs with that opportunity. From the sample in this study, it has been found that big data is not as widely adopted as initially thought. Organisations are still finding value in the traditional structured data around which they have built their business. This, in turn, has provided little opportunity for traditional DBAs to move into a big data environment within South Africa, and has potentially resulted in DBAs also looking to pursue other IT careers.

This research will thus assist organisations and educationalists in defining a potential career path for DBAs towards a role within big data. It will also help educators in designing curricular that can adapt the DBAs skill set towards one of these job families.

7.6.1 Further Work

Further research should be considered in the following areas:

(i) This study has found that the rate of adoption of big data within organisations in South Africa is slower than expected. Further research could explore the extent to which organisations across various sectors in South Africa are adopting big data to extract business value, and what is inhibiting them from doing so.

(ii) This study has found that DBAs want to pursue other IT careers and thus further research could explore if this is a general trend amongst DBAs or specific to this study.

(iii) No specific career path has been found in this study to develop staff for big data roles. Further research could be performed to investigate how industry and educationalists could address this.

7.7 Research Limitations

(i) This study was exploratory. The research methodology and techniques subsequently applied were not aimed at generalising the findings to the rest of the DBA population. Resources to conduct the study across more industries within South Africa were also constrained.

(ii) Because organisations consider their data as a valuable commodity, and because DBAs have access such data, there was hesitancy from some organisations and individuals to participate. The fear of a potential breach of IP or sensitive information was of concern resulting in difficulties finding interviewees willing to participate in the research.

(iii) The extent by which the findings of this study could identify skills actually being employed within a big data environment was constrained through the slow adoption of big data in the sample population. A study should be done to identify the extent to which different technologies are currently being deployed within a big data environment.

CHAPTER EIGHT: MY RESEARCH JOURNEY

Now that I have reached the end of my research journey, I would like to reflect on the road travelled to reach this destination. While writing these few words in conclusion of my thesis, I am relieved that I was able to come this far. It was rewarding to experience the rigour with which research must be conducted, yet it was stressful and frustrating at times.

The initial research methodology for this study was a case study using an online questionnaire as my data collection instrument. Having identified and been given access to a sample of in excess of 50 DBAs, I was very excited. However, at the time of data collection, the sample had shrunk to a size not suitable to conduct a reasonable statistical analysis. I then needed to reach out to other organisations to increase my sample to get a reasonable response rate. This was a frustrating process. Cold calling many organisations either resulted in a flat-out “no” or they never responded to my emails. I tried more than 25 organisations, every time getting excited at the potential opportunity but eventually ending in disappointment. I got the sense that many organisations did not feel comfortable with the idea of granting access to DBAs, which being a DBA myself, I could understand.

Getting a foot in the door at an organisation was a time-consuming process and needed patience. While it was promising upon receiving verbal consent from many organisations, the majority did not materialise. It was required that I get consent from the organisation’s Chief Information Officer (CIO). This resulted in a slow turn-around time as individuals in those positions are extremely busy. Only where I had a contact person within the organisation was the turnaround time quicker. This really gave me the sense that many organisations are saturated with requests for research, making it difficult to gain access. Eventually I got access to DBAs within five more organisations. Some organisations allowed me to contact DBAs directly through email, but others sent the request on my behalf. Even though I received consent from an organisational level, getting DBAs to take part was even harder.

While I continued the search for more organisations to avail their DBAs to this study, I started the data collection process. I gave respondents two weeks to respond to my initial invitation to participate. Thereafter I sent a follow-up email. This was another slow process as I was at the mercy of those who voluntarily would agree to participate. While my supervisor was absent due to being ill, my co-supervisor had to decide on an alternate data collection instrument as I still did not have enough

respondents after all my endeavours. It was decided that I shall conduct interviews, semi-structured interviews, and this was interesting. While some individuals were willing to participate through the anonymity provided to them by the online questionnaires, they were less enthusiastic to participate in face-to-face interviews.

Those who were willing to take part I met at various locations and times to do the interviews which were convenient for them. This ranged from their office, coffee shops and even fast food chains. The challenging part of meeting in coffee shops and fast food chains was the background noise. I had no control over the noise and made transcribing the interviews a difficult process. I was eventually given an opportunity to get access to a sufficiently large enough sample of DBAs, but I had to travel to Johannesburg. I booked my ticket as soon as I could, and I was fortunate to get some rich data.

There were many other distractions during my time of studying. Student protests, issues with my registration all prevented me from making contact and spending time with my supervisor for lengthy periods. I was also required to take a leave of absence for more than a year to complete my Oracle DBA certifications. This severely delayed my progress, and I was required to revisit the literature I had previously researched.

Even with all the challenges I encountered, I am glad that I was able to persist. I could understand why respondents were tentative, being a DBA myself, I would most probably also have been, who knows. But when the opportunity eventually came about to conduct the interviews, I could relate with many of the respondents in this study. It seemed all too familiar, my daily experiences and also having transitioned from a software development role to that of a DBA. But with all the challenges, the support and encouragement I got from those around me gave me strength to push through. It made the whole experience worthwhile and rewarding.

REFERENCES

Aasheim, C.L., Williams, S., Rutner, P. & Gardiner, A. 2015. Data analytics vs. data science: a study of similarities and differences in undergraduate programs based on course descriptions. *Journal of Information Systems Education*, 26(2):103-115.

Abadi, D. *et al.* 2016. The Beckman report on database research. *Communications of the ACM*, 59(2):92-99.

Adriana, J. & Holanda, M. 2016. Brasilia's database administrators. *Journal of Information Systems Engineering & Management*, 1(3):149-157.

Ahmed, F. F. 2015. Comparative analysis for cloud based e-learning. *Procedia Computer Science*, 65:368-376.

Alguliyev, R. M., Gasimova, R. T. & Abbasli, R. N. 2017. The obstacles in big data process. *International Journal of Modern Education and Computer Science*, 9(3):28.

Almeida, F. & Calistru, C. 2013. The main challenges and issues of big data management. *International Journal of Research Studies in Computing*, 2(1):11-20.

Almeida, F. L. F. 2017. Benefits, challenges and tools of big data management. *Journal of Systems Integration*, 8(4):12-20.

Arnold, J., Glavic, B. & Raicu, I. 2015. HRDBMS: a NewSQL database for analytics. *Proceedings.* The 2015 IEEE International Conference on Cluster Computing (CLUSTER), Chicago, IL, USA, 8-11 September 2015. IEEE: 519-520.

Arora, I. & Gupta, A. 2012. Cloud databases: a paradigm shift in databases. *International Journal of Computer Science Issues (IJCSI)*, 9(4):77-83.

Asthana, P. 2011. Big data and little data. *Forbes.com*. http://i.dell.com/sites/doccontent/business/solutions/power/en/documents/ps4q11-20120123-recentcitings.pdf [17 June 2013].

Atieno, O. P. 2009. An analysis of the strengths and limitation of qualitative and quantitative research paradigms. *Problems of Education in the 21st Century*, 13(1):13-18.

Atzeni, P., Jensen, C.S., Orsi, G., Ram, S., Tanca, L. & Torlone, R. 2013. The relational model is dead, SQL is dead, and I don't feel so good myself. *ACM SIGMOD Record*, 42(2):64-68.

Australian Government Department of Finance and Deregulation. 2013. *Big data strategy – issues paper*. https://www.finance.gov.au/files/2013/03/big-data-strategy-issues-paper1.pdf [23 April 2013].

Bailey, J. L. 2014. Non-technical skills for success in a technical world. *International Journal of Business and Social Science*, 5(4):1-10.

Benbasat, I., Goldstein, D. K. & Mead, M. 1987. The case research strategy in studies of information system. *MIS Quarterly*: 369-386.

Binani, S., Gutti, A. & Upadhyay, S. 2016. SQL vs. NoSQL vs. NewSQL - a comparative study. *Communications on Applied Electronics (CAE)*, 6(1):43-46.

Bizer, C., Boncz, P., Brodie, M.L. & Erling, O. 2012. The meaningful use of big data: four perspectives - four challenges. *ACM SIGMOD Record*, 40(4):56-60.

Bloem, J., van Doorn, M., Duivestein, S., van Manen, T. & van Ommeren, E. 2012. Creating clarity with big data. *Sogeti VINT, 2012.*

Blumberg, R. & Atre, S. 2003. The problem with unstructured data. *DM Review*, 13:42-46.

Boost your career as a database administrator. 2013. *Athena Information Solutions Pvt. Ltd.* New Dehli.

Borkar, V. R., Carey, M. J. & Li, C. 2012a. Big data platforms: what's next?. *XRDS: Crossroads, The ACM Magazine for Students*, 19(1):44-49.

Borkar, V.R., Carey, M.J. & Li, C., 2012b. Inside “big data management”: ogres, onions, or parfaits?. *Proceedings.* The 15th international conference on extending database technology, Berlin, Germany, 27-30 March 2012. ACM: 3-14.

Botta, A., De Donato, W., Persico, V. & Pescapé, A. 2014. On the integration of cloud computing and internet of things. *Proceedings.* The 2014 International Conference on Future Internet of Things and Cloud*,* 27-29 August 2014. IEEE Computer Society: 23-30.

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

Cao, L. 2016. Data science: nature and pitfalls. *IEEE Intelligent Systems*, (5):66-75.

Corbellini, A., Mateos, C., Zunino, A., Godoy, D. & Schiaffino, S. 2017. Persisting big-data: the NoSQL landscape. *Information Systems*, 63:1-23.

Costa, C. & Santos, M. Y. 2017. Big data: state-of-the-art concepts, techniques, technologies, modeling approaches and research challenges. *IAENG International Journal of Computer Science*, 44(3):285-301.

Courtney, M. 2012. The larging-up of big data. *Engineering & Technology*, 7(8):72-75.

Courtney, M. 2013. Puzzling out big data. *Engineering & Technology*, 7(12):56-60.

Crossan, F. 2003. Research philosophy: towards an understanding. *Nurse Researcher (through 2013)*, 11(1):46.

Davenport, T. H., Barth, P. & Bean, R. 2012. How ‘big data’ is different. *MIT Sloan Management Review*: 54(1).

Davenport, T.H. & Patil, D.J. 2012. Data scientist: the sexiest job of the 21st century. *Harvard Business Review*: 70-76, October.

De Mauro, A., Greco, M., Grimaldi, M. & Ritala, P. 2017. Human resources for big data professions: a systematic classification of job roles and required skill sets. *Information Processing and Management*, 54(5):807-817.

Dearnley, C. 2005. A reflection on the use of semi structured interviews. *Nurse Researcher*: 13(1).

Debortoli, S., Müller, O. & vom Brocke, J. 2014. Comparing business intelligence and big data skills: a text mining study using job advertisements. *Business and Information Systems Engineering*, 6(5):289-300.

DiCicco-Bloom, B. & Crabtree, B. F. 2006. The qualitative research interview. *Medical Education*, 40(4):314-321.

Diebold, F. 2003. Big data dynamic factor models for macroeconomic measurement and forecasting. (Discussion of Reichlin and Watson papers), In M. Dewatripont, L.P. Hansen and S. Turnovsky (Eds.), Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society. Cambridge: Cambridge University Press: 115-122.

Dijcks, J-P. 2011. *Oracle: big data for the enterprise.* An Oracle White Paper. Oracle Corporation.

Ellis, T. J. & Levy, Y. 2009. Towards a guide for novice researchers on research methodology: review and proposed methods. *Issues in Informing Science and Information Technology*, 6:323-337.

Erraissi, A., Belangour, A. & Tragha, A. 2017. Digging into hadoop-based big data architectures. *International Journal of Computer Science Issues (IJCSI)*, 14(6):52-59.

Feuerlicht, G. 2010. Database trends and directions: current challenges and opportunities. *DATESO*: 163-174.

Galbraith, J. R. 2012. The evolution of enterprise organization designs. *Journal of Organization Design*, 1(2):1-13.

Gray, D. E. 2013. Theoretical perspectives and research methodologies. *Doing research in the real world*. *Sage*: 16-38.

Gupta, D. & Siddiqui, S. 2013. Understanding the power of big data. *International Journal of Research Review in Engineering Science and Technology (IJRREST)*, 2(1):173-178.

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A. & Khan, S.U. 2015. The rise of “big data” on cloud computing: review and open research issues. *Information Systems. Elsevier*, 47:98-115.

Hecht, R. & Jablonski, S. 2011. NoSQL evaluation: a use case oriented survey. *Proceedings.* The 2011 International Conference on Cloud and Service Computing, Washington, DC, USA, 12-14 December 2011. IEEE Computer Society: 336-341.

Hitzler, P. & Janowicz, K. 2013. Linked data, big data, and the 4th paradigm. *Semantic Web*, 4(3):233-235.

Holden, M.T. & Lynch, P. 2004. Choosing the appropriate methodology: understanding research philosophy. *The Marketing Review*, 4(4):397-409.

Holt, V., Ramage, M., Kear, K. & Heap, N. 2015. The usage of best practices and procedures in the database community. *Information Systems*, 49:163-181.

Hu, H., Wen, Y., Chua, T.S. & Li, X. 2014. Toward scalable systems for big data analytics: a technology tutorial. *IEEE Access*, 2:652-687.

Hunter, P. 2013. Journey to the centre of big data. *Engineering & Technology*, 8(3):56-59.

Jacobs, A. 2009. The pathologies of big data. *Communications of the ACM*, 52(8):36-44.

Joe, K., Raben, F. & Phillips, A. 2016. The ethical issues of survey and market research. In Wolf, C., Joye, D., Smith, TW & Fu, Y. *The SAGE Handbook of Survey Methodology*: 77-86.

Johnston, L.G. & Sabin, K. 2010. Sampling hard-to-reach populations with respondent driven sampling. *Methodological Innovations Online*, 5(2):38-48.

Jupp, V. 2006. *The sage dictionary of social research methods*. SAGE Publications, Ltd, London.

Kaisler, S., Armour, F., Espinosa, J.A. & Money, W. 2013. Big data: issues and challenges moving forward. *Proceedings.* The 2013 46th Hawaii International Conference on System Sciences, Washington, DC, USA, 7-10 January 2013. IEEE Computer Society: 995-1004.

Kaplan, B. & Maxwell, J.A. 2005. Qualitative research methods for evaluating computer information systems. *Evaluating the Organizational Impact of Healthcare Information Systems*: 30-55.

Khan, N., Yaqoob, I., Hashem, I.A.T., Inayat, Z., Ali, W.K.M., Alam, M., Shiraz, M. & Gani, A. 2014. Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal, 2014*: 1-18.

King, E. 2015. The real world of the database administrator. *Unisphere Research*. http://www.dbta.com/dbta-downloads/researchreports/the-real-world-of-the-database-administrator-5237.aspx [25 August 2018].

King, E. 2017. DBAs face new challenges: trends in database administration. *Unisphere Research*. http://www.dbta.com/dbta-downloads/whitepapers/dbas-face-new-challenges-trend-in-database-administration-7650.aspx [25 August 2018].

Kumar, R. & Charu, S. 2014. NewSQL databases: scalable RDBMS for OLTP needs to handle big data. *International Journal of Modern Computer Science (IJMCS)*, 3(1):13-17.

Kvale, S. 2011a. Conducting an interview. In *Doing Interviews, Qualitative Research Kit.* SAGE Publications, Ltd. London: 52-66.

Kvale, S. 2011b. Introduction to interview research. In *Doing Interviews, Qualitative Research Kit.* SAGE Publications, Ltd. London: 2-10.

Kvale, S. 2011c. Planning an interview study. In *Doing Interviews, Qualitative Research Kit.* SAGE Publications, Ltd. London: 34-50.

Kvale, S. 2011d. Transcribing interviews. In *Doing Interviews, Qualitative Research Kit.* SAGE Publications, Ltd. London: 93–100.

Leavitt, N. 2010. Will NoSQL databases live up to their promise?. *Computer*: 43(2).

Lee, K. K. Y., Tang, W. C. & Choi, K. S. 2013. Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage. *Computer Methods and Programs in Biomedicine,* 110(1):99-109.

Litecky, C., Aken, A., Ahmad, A. & Nelson, H.J. 2010. Mining for computing jobs. *IEEE Software*, 27(1):78-85.

Liu, K. & Dong, L. 2012. Research on cloud data storage technology and its architecture implementation. *Procedia Engineering*, 29:133–137.

Louise Barriball, K. & While, A. 1994. Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing*, 19(2):328-335.

Luna, D.R., Mayan, J.C., García, M.J., Almerares, A.A. & Househ, M. 2014. Challenges and potential solutions for big data implementations in developing countries. *IMIA Yearbook of Medical Informatics*, 9:36-41.

Madden, S. 2012. From databases to big data. *IEEE Internet Computing*, 16(3):4-6.

Madison, M., Barnhill, M., Napier, C. & Godin, J. 2015. NoSQL database technologies. *Journal of International Technology and Information Management*, 24(1):1-14.

Maguire, M. & Delahunt, B. 2017. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education*: 9(3).

Mai, P.T.A., Nurminen, J.K. & Di Francesco, M. 2014. Cloud databases for internet-of-things data. *Proceedings.* The 2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom), Taipei, Taiwan, 1-3 September 2014. IEEE: 117-124.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. 2011. Big data: the next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation [18 April 2013].

Marshall, M. N. 1996. Sampling for qualitative research. *Family Practice*, 13(6):522-526.

Matthews, R. J. 2017. A theory for everything? Is a knowledge of career development theory necessary to understand career decision making? *European Scientific Journal (ESJ)*, 13(7):320-334.

Mays, N. & Pope, C. 1995. Qualitative research: rigour and qualitative research. *BMJ*, 311(6997):109-112.

McAfee, A. & Brynjolfsson, E. 2012. Big data: the management revolution. *Harvard Business Review*.

McGehee, B. M. 2009. *How to become an exceptional DBA*. 2nd ed. Cambridge, UK: Simple-Talk Publishing.

McKendrick, J. 2014. The vanishing database administrator: survey of data professionals' career aspirations. *Unisphere Research*. http://www.dbta.com/dbta-downloads/researchreports/the-vanishing-database-administrator-survey-of-data-professionals-career-aspirations-4791.aspx [9 September 2018].

Mishra, R. & Sharma, R. 2015. Big data: opportunities and challenges. *International Journal of Computer Science and Mobile Computing*, 4(6):27-35.

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 and Application*: 6(4):1-13.

Moore, P. 2010. Database administrator. *Tech Directions*: 24-25.

Morgan, G. & Smircich, L. 1980. The case for qualitative research. *Academy of Management Review*, 5(4):491-500.

Murawski, M. & Bick, M. 2017. Demanded and imparted big data competences: towards an integrative analysis. *Proceedings*. The 25th European Conference on Information Systems (ECIS), Guimarães, Portugal, 5-10 June 2017: 1375-1390.

Myers, M. D. 1997. Qualitative research in information systems. *MIS Quarterly*, 21(2):241-242, June 1997. *MISQ Discovery*, archival version, June 1997, http://www.misq.org/supplements/. *Association for Information Systems (AISWorld) Section on Qualitative Research in Information Systems*, updated version, last modified: 4 September 2018. www.qual.auckland.ac.nz [22 November 2018].

Nayak, A., Poriya, A. & Poojary, D. 2013. Type of NoSQL databases and its comparison with relational databases. *International Journal of Applied Information Systems*, 5(4):16-19.

NBD-PWG. 2015. NIST big data interoperability framework: volume 6, reference architecture. *NIST Special Publication 1500-6*, 6:1-62.

Nizam, T. & Hassan, S.I. 2017. Big data: a survey paper on big data innovation and its technology. *International Journal of Advanced Research in Computer Science*, 8(5):2173-2177.

Orb, A., Eisenhauer, L. & Wynaden, D. 2001. Ethics in qualitative research. *Journal of Nursing Scholarship*, 33(1):93-96.

Oussous, A., Benjelloun, F.Z., Lahcen, A.A. & Belfkih, S. 2018. Big data technologies: a survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4):431-448.

Oxford Dictionaries | English. 2018. *English dictionary, thesaurus, & grammar help | Oxford dictionaries.* https://en.oxforddictionaries.com [15 October 2018].

Peshkar, P. & Ghosekar, P. 2015. Role of database administrator in the I.T. industry. *International Journal of Emerging Research in Management & Technology*, 4(12):41-44.

Pingle, Y., Kohli, V., Kamat, S. & Poladia, N. 2012. Big data processing using apache hadoop in cloud system. *Proceedings.* The National Conference on Emerging Trends in Engineering & Technology: 475-479.

Pokorný, J. 2013. New database architectures: steps towards big data processing. *Proceedings.* The IADIS European Conference Data Mining 2013, Prague, Czech Republic, 22-24 July 2013: 3-10.

Polakowski, M. 2009. Database administrators : it's all in a day's work. *Certification Magazine*, 11(12):31-33.

Priyanka & AmitPal. 2016. A review of NoSQL databases, types and comparison with relational database. *International Journal of Engineering Science and Computing*, 6(5):4963-4966.

Provost, F. & Fawcett, T. 2013. Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1):51-59.

Purcell, B.M. 2014. Big data using cloud computing. *Journal of Technology Research*, 5:1-8.

Raut, A. B. 2017. NoSQL database and its comparison with RDBMS. *International Journal of Computational Intelligence Research*, 13(7):1645-1651.

Remenyi, D., Pather, S. & Klopper, R. 2011. Some philosophical assumptions underpinning academic research. *Management, Informatics and Research Design*: 354-373.

Ridge, M., Johnston, K.A. & O'Donovan, B. 2015. The use of big data analytics in the retail industries in South Africa. *African Journal of Business Management*, 9(19):688-703.

Ryan, G. W. & Bernard, H. R. 2003. Techniques to identify themes. *Field Methods*, 15(1):85-109.

Seale, C. & Silverman, D. 1997. Ensuring rigour in qualtative research. *The European Journal of Public Health*, 7(4):379-384.

Sharma, S. 2013. Big data landscape. *International Journal of Scientific and Research Publications*, 3(6):1-8.

Siddiqa, A., Karim, A. & Gani, A. 2017. Big data storage technologies: a survey. *Frontiers of Information Technology & Electronic Engineering*, 18(8):1040-1070.

Silva, Y. N., Almeida, I. & Queiroz, M. 2016. SQL: from traditional databases to big data. *Proceedings.* The 47th ACM Technical Symposium on Computing Science Education, Memphis, TN, USA, 2-5 March 2016. ACM: 413-418.

Simmonds, T. 2013. Teaching database administration in the world of big data and small budgets. *Proceedings.* The 11th International Workshop on the Teaching, Learning and Assessment of Database, University of Sunderland, 5 July 2013: 45-51.

Simplilearn.com. 2018. *Apache cassandra certification training.* https://www.simplilearn.com/big-data-and-analytics/apache-cassandra-certification-training [31 October 2018].

Sowmya, M. & Sravanthi, N. 2017. Big data: an overview of features, tools, techniques and applications. *International Journal of Engineering Science and Computing*, 7(6):13644-13647.

Stake, R.E. 2013. Qualitative research and case study. *Silpakorn Educational Research Journal*, 3(1-2):7-13.

Stonebraker, M., Abadi, D., DeWitt, D.J., Madden, S., Paulson, E., Pavlo, A. & Rasin, A. 2010. MapReduce and parallel DBMSs: friends or foes?. *Communications of the ACM*, 53(1):64-71.

Storey, V. C. & Song, I.Y. 2017. Big data technologies and management: what conceptual modeling can do. *Data & Knowledge Engineering*, 108:50-67.

Sukumar, S.R., Olama, M.M., McNair, A.W. & Nutaro, J.J. 2013. Concept of operations for knowledge discovery from “big data” across enterprise data warehouses. In *Next-Generation Analyst.* International Society for Optics and Photonics, 8758:875805.

Talia, D. 2013. Clouds for scalable big data analytics. *Computer,* 46(5):98-101.

Tauro, C.J., Aravindh, S. & Shreeharsha, A.B. 2012. Comparative study of the new generation, agile, scalable, high performance NoSQL databases. *International Journal of Computer Applications*, 48(20):1-4.

Thomas, D. 2003. A general inductive approach for qualitative data analysis. University of Auckland, New Zealand. http://www.health.auckland.ac.nz/hrmas/resources/inductive2003.pdf [24 August 2013].

Tittel, E. 2003. Certified expert: working as a database administrator. *Certification Magazine*: 44.

Tuli, F. 2010. The basis of distinction between qualitative and quantitative research in social science: reflection on ontological, epistemological and methodological perspectives. *Ethiopian Journal of Education and Sciences*, 6(1):97-108.

Turner III, D. W. 2010. Qualitative interview design: a practical guide for novice investigators. *The Qualitative Report*, 15(3):754-760.

University.mongodb.com. 2018. *MongoDB professional certification.* https://university.mongodb.com/certification [31 October 2018].

Vaismoradi, M., Turunen, H. & Bondas, T. 2013. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15(3):398-405.

Vemuganti, G. 2013. Metadata management in big data. *Big Data: Countering Tomorrow’s Challenges*, 11(1):3-9.

Volk, M., Bosse, S. & Turowski, K. 2017. Providing clarity on big data technologies: a structured literature review. *Proceedings.* The 2017 IEEE 19th Conference on Business Informatics, Thessaloniki, Greece, 24-27 July 2017. IEEE: 388-397.

Wahyuni, D. 2012. The research design maze: understanding paradigms, cases, methods and methodologies. *Journal of Applied Management Accounting Research*, 10(1):69-80, Winter 2012.

Walsham, G. 1995. Interpretive case studies in IS research: nature and method. *European Journal of Information Systems*, 4(2):74-81.

Walsham, G. 2006. Doing interpretive research. *European Journal of Information Systems*, 15(3):320-330.

Yin, R. K. 2006. Mixed methods research: are the methods genuinely integrated or merely parallel. *Research in the Schools*, 13(1):41-47.

Zaki, A. K. 2014. NoSQL databases: new millennium database for big data, big users, cloud computing and its security challenges. *International Journal of Research in Engineering and Technology (IJRET)*, 3(15):403-409.

Zanoon, N., Al-Haj, A. & Khwaldeh, S. M. 2017. Cloud computing and big data is there a relation between the two: a study. *International Journal of Applied Engineering Research*, 12(17):6970-6982.

Zaslavsky, A., Perera, C. & Georgakopoulos, D. 2013. Sensing as a service and big data. *arXiv preprint arXiv:1301.0159.*

Zhu, T., Xiao, S., Zhang, Q., Gu, Y., Yi, P. & Li, Y. 2015. Emergent technologies in big data sensing: a survey. *International Journal of Distributed Sensor Networks*, 11(10):902982.

APPENDIX A: University consent template

<<On company letterhead>>

I <<*insert name*>>, in my capacity as <<*insert position in company*) at <<*insert company name*>> give consent in principle to allow <<*insert student name*>>, a student at the Cape Peninsula University of Technology, to collect data in this company as part of his/her M Tech (IT) research. The student has explained to me the nature of his/her research and the nature of the data to be collected.

This consent in no way commits any individual staff member to participate in the research, and it is expected that the student will get explicit consent from any participants. I reserve the right to withdraw this permission at some future time.

In addition, the company's name may or may not be used as indicated below. (Tick as appropriate.)

	Thesis	Conference paper	Journal article	Research poster
Yes				
No				

_____ _____

<<*Insert name*>> <<*Insert date*>>

APPENDIX B: Interview consent form

Cape Peninsula University of Technology

Interview Consent Form

I, [redacted], in my private capacity give consent in principle to allow Romney Davids (Student Number: 195064720), a student at the Cape Peninsula University of Technology, to conduct a face-to-face interview with me as part of his MTech (IT) research. The student explained to me the nature of his research and the nature of the data to be collected.

I reserve the right to withdraw this permission at any time during the interview.

In addition, my name may or may not be used as indicated below. (Tick as appropriate.)

	Thesis	Conference paper	Journal article	Research poster
Yes				
No	No	No	No	No

[redacted]

Printed Name

28/02/2018

Date

[redacted]

Participant's Signature

APPENDIX C: Interview guide

Semi-structured-questionnaires

Interview schedule:

Introductory remarks: There is currently a shortage of specialists to administer big data environments in organisations. Even though DBAs manage the organisations traditional RDBMS, they do not have the required skills needed to work in a big data environment. Organisations might therefore not be able to extract the potential value that big data holds, if they do not have the necessary skilled staff to manage and analyse big data within the organisation.

The aim: This study seeks to explore the difference in skillset between a DBA administering a RDBMS and similar position in a big data environment.

We are kindly requesting answers to the questions listed below in your good faith. Your answers will be used specifically for this study purposes only and they will be treated with the highest degree of confidentiality and privacy. Also, participation in this interview is voluntary and allows anonymity as well as autonomy.

Section A: participant's details

Name: _____ surname:_____ Position: _____	Date:_____ Contact No:_____

Section B: Questions

Guidelines to answering the interview questions:

- **Prompt Questions** – these questions serve merely as a guide to answering the interview questions

RESEARCH QUESTION	**What is the skills gap between the DBA working in a RDBMS environment and big data specialist?**	
RESEARCH SUB QUESTION 1	**What is the current skillset and responsibilities of a DBA working in a RDBMS environment?**	
Interview Question	**Prompt Questions**	**Interviewee Response**
(1.1) Tell me about yourself?	(1) Please provide a brief history of your background as an I.T. professional (2) What and where did you study? (formal qualifications) (3) How did you decide to become a DBA? Is it something you studied for or something that developed as time went by?	
(1.2) What does your typical day as a DBA look like?	(1) How do you start your day? (2) What are your primary responsibilities as a DBA? (3) Are there specific functions/responsibilities you are required to perform daily? (4) Are there any other non-DBA related functions you perform? (5) Are you required to interface with other teams within your organization or with external customers?	
(1.3) What are the core skills you would say is important to have to be a DBA?	(1) What are the different areas a DBA should have knowledge of or have an understanding of? (2) What are the current technologies you are exposed to? (3) How do you keep your skillset current with technologies constantly changing and newer technologies being introduced?	
(1.4) What other skills / competencies should a DBA have that one would not directly associate with being a DBA?	(1) Should a DBA have programming / analytical skills / security etc.? (2) Non-technical skills such as organizational / communication / interpersonal / time-management / adapt to new tech / problem-solving / team-work etc.	
(1.5) What are the essential qualifications / training required to become a DBA?	(1) How important is a formal qualification? (2) What role does vocational training / vendor certifications have within the DBA profession? (3) How does vocational training / vendor certifications contribute towards the DBAs skills? (4) How do you go about updating your skillset?	

RESEARCH SUB QUESTION 2	What are the skills and technology requirements of a big data environment?

Interview Question	Prompt Questions	Interviewee Response
(2.1) What are your thoughts on Big Data?	(1) Is it something you foresee yourself becoming involved with in your career? (2) What do you think a typical day of a Big Data DBA will look like?	
(2.2) How do you see Big Data affecting the way that DBAs in the future?	(1) Would it change the way DBAs work? (2) Would it have an impact on the DBAs responsibilities as we know it today within the RDBMS environment? (3) Would there be other requirements placed on the shoulders of the DBA such as security etc.?	
(2.3) What are your thoughts of the skills required in a Big Data environment?	(1) Would there be a requirement to formally re-skill oneself? (2) Do traditional DBAs already have sufficient skills to manage a Big Data environment? (3) Would there be a need to possibly re-certify yourself and if so, what type of certifications will be required? (4) What role does vocational training / vendor certifications have within the Big Data environment? (5) Would there be any other skillset required in addition to what is currently expected of traditional DBAs? (coding / analytical / data mining / BI etc.)	
(2.4) What are your thoughts of the various technologies emerging within the Big Data landscape?	(1) Will these emerging technologies future be short-lived or will this be the way forward for DBAs? (2) Have you started looking at technologies & competencies required within the Big Data environment? (3) What should DBAs do to stay current with the emergence of all these new technologies & trends?	

RESEARCH SUB QUESTION 3	What career plans do organisations have for DBAs with regards to big data management?

Interview Question	Prompt Questions	Interviewee Response
(3.1) What are the current career development plans or paths that are in place for DBA's within your organization?	(1) Would DBAs be able to pursue a career within Big Data? (2) What career pathing has been outlined for DBAs with regard Big Data management? (3) How would you describe your prospects towards moving to a role within a Big Data environment? (4) Where do you see yourself within the next 5-10 years? Still working as a DBA?	
(3.2) What can organisations do to help DBAs to get the needed skills to work in a Big Data environment?	(1) Do you think this is dependent on the direction the company has defined for itself in terms of its technology stack? (2) How will the adoption of open-source technologies that will assist DBAs in moving towards a role within a Big Data environment?	

Thank you for your time and patience in answering the questions. Your contribution is highly appreciated.

APPENDIX D: Interview transcript for two respondents

(The transcripts and audio recordings for all respondents available upon request)

Interview Transcript #1

Interviewer name	Romney Davids (RD)	**Interview date**	Thurs, 2017/09/14
Interview start date	12:30	**Interview duration**	24 min 27 sec
Interviewee name	Anonymous (R1)	**Audio filename**	R01.1.20170914
Interviewee occupation	Database Administrator Lead	**Pseudonym**	R1
Notes	Any company names or individual's names have been edited out of the interview transcript.		

RD Recording interview with respondent 1. Do you grant me permission to record this interview?

R1 I do.

RD Thank you

RSQ1: What is the current skillset and responsibilities of a DBA working in a RDBMS environment?

Q1.1 ***Tell me a bit about yourself. Just a quick history of your IT background***

R1 OK. Well my position is Team Lead DBA at *[CompanyNameWithheld]*. I've been a DBA for more than 10 years. Probably about 14 years. I started off as a Oracle Developer. So, I've worked from a development background and I have transitioned to a DBA about 12years ago.

RD Okay …

R1 I've always been involved in Oracle and being a DBA

RD And is there something that has made you decide to become a DBA or is it something you always wanted to do?

R1 (Hmm) …, so I started off as a developer and when I was working as a developer, the DBA guys always seemed to be very knowledgeable and well paid, so I always thought it was a good position to aim for

RD And what did you study to become a developer? Did you specifically study to become a developer or is it just …

R1 No, I initially studied computer science. Um but I did not finish my degree I still had a couple of credits outstanding when I completed so then I ja, I didn't specifically study to become a developer

RD OK. That's interesting …

Q1.2 ***Being a DBA, what does your typical day look like?***

R1 (Hmm) …

RD Is there specific routines you use to start your day or …

R1 Well ja, usually in the mornings I go through my emails to see if there are any messages from our alert system. We have a system that checks alert logs and makes sure that it does health checks on all of our databases. So I usually go through those and check to make sure that the databases are up

and running and if there are any problems then I will resolve those so that's usually the beginning part of my day and the rest of my day is filled with meetings and then attending to other work items, things we have to do for, regarding ja….

RD Is there any specific responsibilities that you can define as being primary responsibilities of a DBA?

R1 Well we primary responsible for providing database services to the floor. I our case that's developers. Developers and testers. So our primary responsibility is to make sure that those databases are up available … (hmm) … and I guess to keep those databases available as much as possible because they all depend on it for their work

RD Okay. Are there any non-DBA related functions that you are required to perform? Outside of the general scope of a DBA or that's defined for a DBA?

R1 (Hmm) …, well as a team lead I guess there's a few management responsibilities that are not DBA related. Then we also in our team we've taken responsibility for some systems that are not DBA related. Specifically, the build system and ja, I suppose specifically the build system I can think of so it's a system that's not related to the DBA field but we have taken on and we support it.

RD And is there a requirement to interface with other teams in the organization?

R1 Ja, constantly

Q1.3 ***In terms of core skills for a DBA, are there any specific core skills that you is a requirement for somebody to become a DBA?***

R1 (Hmm) …

RD … or that one should have?

R1 Well, I guess an overall understanding of IT might be a good start. Then specifically you would need, you need some kind of OS skills, for us most of our work is done in Linux or Unix so those, the OS skills are important. Then kinda getting more specific SQL skills, you need to be able to manipulate databases using SQL, that's how you talk to it. And then there are more specific skills would be database management skills, what's involved in being a DBA, all the admin kinds of stuff that a DBA needs to know.

RD Great! Specific technologies you exposed to at the moment?

R1 In regard to? Just in the database side or…

RD Database side or in general as a DBA

R1 Sjo, I can list a lot, well there's Oracle, Linux, Jenkins, Git, Subversion, Java, Windows

RD So it's quite specific to a DBA depending on where he works …

R1 OK, ja, Oracle, mainly Oracle is the main technology

RD In terms of your current skills set how do you keep that current? How do you keep abreast of the latest technologies?

R1 It's all self, what you call self-directed learning, I mean I use it's kind of part of my daily routine is to stay up to date, read the Oracle blogs, read the new features coming out and play around with new features, we are RnD so I do spend quite a lot of time when new versions come out installing that and getting to know that. But it's all self-directed

RD So it's a self-taught

R1 Ja, there are courses and that kind of stuff but that usually comes after the fact, the certification

Q1.4 ***Are there any other skills or competencies that are not directly related to being a DBA that you should have?***

R1 You mean non-technical, soft skills …

RD Soft skills, non-technical skills or other technical skills

R1 I think you got to have very good problem solving skills cause lot of the day is solving problems. (Hmm) … you got to be self-motivated and you got to be, I forgotten the word I was looking for …

RD Communication skills?

R1 Ja, you need good communication skills, good team-work skills because you working with a lot of people not just in the DBA team but across the floor. (Hmm) … I think that's all I can think of. Sorry I'm just having a, there one word, if I remember I will let you know …

Q1.5 ***Okay … and essential qualifications, is there specific training that one requires to become a DBA?***

R1 My opinion or I think the industry standard is you would get a Batchelors Degree or some kind of degree tertiary qualification in Computer Science or Computer studies or whatever you call it these days and then that's usually the entry level and then do some kind of Oracle certification, I think is also most jobs require or would prefer Oracle certification

RD And in terms of that vocational training or vendor certifications, what role does it play, how does it add to the DBA skills

R1 Well I guess the main thing is it adds credibility to what the DBA claims, the skills the DBA has, the fact that he has skills, he is certified, it gives him some kind of credibility. I believe that you get more, your skills get more from on the job, day to day work, than going on a one week course

RD Okay …

R1 That's in my experience, but it also sometimes courses are good because they cover gaps that you may have…

RD So it's a forced learning almost …?

R1 Not forced *(laughter from R1)*, it's on the job, ja, it's forced learning and for any learning to be affective you need to apply it and most people don't apply what they've learned on course

RSQ2: What are the skills and technology requirements of a big data environment?

Q2.1 ***The second part of this will deal with the big data environment. What are your thoughts on big data in general. Is it something you see yourself becoming involved in in your career?***

R1 Well, I don't know if I will specifically become involved in big data but again there's that whole definition of what big data is. Its taken a while, it is becoming clearer what people mean when they

say big data. I think in our job we do handle large volumes of data but whether, it doesn't really fit in with what the definition of what big data is, so no I don't see myself going with what is now becoming known as big data unless I move to another company where, that's big data…

RD Do you perhaps have any idea of what a big data DBAs day will look like or do you think it will be much the same.

R1 I would guess fairly similar although the one difference – one of the differences we development DBAs whereas big data DBAs I imagine are going to be more production focus because big data is about collecting data that's coming in all the time so I guess big data dba will be more production focused and will be having to monitor data feeds, really constantly, because big data, your data feeds don't stop, they just come at you all the time, that's how I understand it, so there'll be more, there'll be less development and more production kind of stuff, monitoring, making sure systems are running that kind of thing – if that makes sense

Q2.2 ***How do you see big data affecting the way DBAs work in the future? I think you probably touched on that a bit in your previous answer.***

R1 Ja, it's a hard question to answer

RD Ok, lets put it this way …

R1 Look I think any DBA going forward, anyone who stays in the DBA field needs to be up to speed with what big data is and maybe big data is going to become less of a side-show and normal database work and big data will become, will kind of merge together. Already you seeing with Oracle they got all these big data connectors and that kind of thing. So big data is just becoming part of normal day-2-day database work. Maybe.

RD And would that affect, that impact on DBAs responsibilities. Would it be any different to what DBAs do in a relational environment?

R1 I don't think so, just again, going back to the incoming data feeds, if you working, it just depends if you in a dev or production kind of environment. I think one of the big things if you become a big data DBA will be taking care of data feeds, data coming in. Because obviously you got to process, big data you got a lot of data to process, so your data feeds can't back up, if you know what I mean – a big hose pipe, it just keeps coming

RD And in terms of specific requirements like security and things like that do you there'll be more focus on that or

R1 I think ja, I think definitely, they are becoming more important in the industry is security and privacy as well. So, but I don't know how that, where privacy fits in a DBAs job because we don't really manage the content of data, we just manage the data from a physical point of view

Q2.3 ***And your thoughts on the skills of a big data DBA? Would you foresee a requirement to reskill?***

R1 You'd need to expand your skillset, obviously the big data tools are different to what we currently use here – so you'd need to learn things like Hadoop for example or whatever the big data – so there will

be a reskilling involved but I don't think, it's not a case of switching skillsets, it's a case of expanding skillset – there's a lot of knowledge which we've learned as DBAs that is very applicable – that's why I say I don't see the day-2-day job being that different just a different set of tools and bigger datasets

RD Do you think the traditional DBA or the RDBMS DBA have sufficient skills to manage this big data environments?

R1 With the right tools I think that they do, ja

RD And would there be a possibility or requirement to re-certify themselves?

R1 Probably advisable, ja – because, if you going, as I say, it gives you the additional credibility when you're out there in the market – so if you do reskill then it make sense to re-certify yourself – otherwise you wasting your money really

RD And would there be any additional skillsets or other skillset required that is currently expected – or is there any other skillset that a big data DBA will be required to have? That's any different to …

R1 … to current ones – You know apart from different tools, I don't think so unless you starting to look at actual data science and data mining and all that kind of stuff – but I'm excluding those things from this – I'm thinking you get more data engineers and you get data – hmm, and that the way it seems to be moving you get data engineers and data scientists – and the data engineers are us, going to be the DBAs – so in terms of skillsets, probably not, I think it's the same kind of job, just different tools and …. I don't think there's a big – as I say the tools are different

Q2.4 ***Your thoughts on the various technologies within the big data landscape – are they, do you think they'll be short-lived or is it the way forward for DBAs?***

R1 I wouldn't, well I haven't and I wouldn't to rush off to become a big data, to drop everything and learn all the big data toolsets – but definitely there are some toolsets that are standing out from the pack like Hadoop and Apache Spark and R I think is one of the other big ones – so there are definitely now becoming – it's becoming clear what the tools are that are going to move forward – so if I was moving that way then definitely there are things you can follow up on – that answer your question

RD Have you started looking those technologies or competencies required for a big data?

R1 I have, I've kind of messed around with Hadoop and but ja, to some degree, just so that I'm familiar what – how it all hangs together – I haven't done any practical kind of work with it no

RD And with all these new technologies emerging, how should DBAs stay current with all the emergence of all these technologies? Would it be much of the same in terms of the way they currently upskilling themselves or keep themselves current or …

R1 I think we still at a stage where DBAs need to decide if they want to become big data DBAs or whether they want to just be RDBMS DBAs – I think there's still a distinction but later on it will merge, so at this point if you decide you want to be a DBA, a big data DBA, then you need to go an move to those skillsets and I think, but if you are currently wanting to stay as a RDBMS DBA I don't think you need to investigate that stuff too much more than knowing what it's about – well that's kind of the approach I've taken

RSQ3: What career plans do organisations have for DBAs with regards to big data management?

Q3.1 ***The last section will deal with career plans and the way forward for DBAs. What are the current career development plans or paths that's in place within your organization for DBAs? Would you be able to pursue a career in big data?***

R1 Not in our current organization – I don't think there's any scope for big data DBAs and there are no career plans

RD And in terms of career pathing has anything been … okay you've …

R1 no, but I can cover it again, so definitely I think at this stage you got to decide whether and I'm not talking about in our company, I think in the industry in general, If you want to go with the big data DBA you need to make that choice and then you need to follow that path - I think the path is independent, not independent but it does split from being a relational database DBA although you could do both

RD How would you describe your prospects of moving towards a role within a big data environment?

R1 If I was going to go that route I think the prospects would be good but it would be a case of specifically going and investigating and reskilling – there would be a reskilling and re-certification involved but I think I would quite easily make the transition – I'm just not familiar with the industry because as of yet I don't think big data industry, and this can be completely wrong, in cape town has grown much

RD Just going back to that question, do you think it's your experience as a traditional DBA that would stand you in good stead to become a big data DBA or is it …

R1 No, definitely - as I say I don't think there's – I think just to kind of sum up what I think where lot of the questions we're answering up – I think someone who's been a traditional DBA could make the move quite easily but they would have to reskill in addition to their current skillset – but anyone who was kind of trying to move forward and trying to decide which – there is at some point, I guess you got to choose whether you – the two paths are split – you know what I'm saying – currently at some point you either remain a traditional DBA with your current skills and you just keep that current or if you decide to become a big data DBA, your prospects are good but you have to take, you have to reskill and re-certify- ja, it's not going to be difficult it's just something that has to be done

RD And where would you see yourself in 5-10 years' time

R1 Retired, no I would probably remain traditional at this point just basically because this company is not going big data in the foreseeable future – that being said, we'll see what the industry is doing as well – at this point, big data is still a bit of a, I mean it's not a buzzword it is becoming real but I don't where the jobs are, you know

Q3.2 ***And is there anything specific organizations can do to help DBAs to get the required skills***

R1 Well it just depends of the organization is going that way but there must be certain training, certifications for things like Cassandra, for Hadoop or Apache Spark – so if a company is, if this

company were to tomorrow have a sudden big data requirement then there are definitely certifications and trainings that they can send people on out there I would imagine because that is now fairly normal thing big data

RD Okay. And lastly the adoption of open-source technologies within the organization, would that assist DBAs in moving that direction or is it still dependent on …

R1 Ja, I think a lot of the, the majority actually if I think of it of the big data stuff is open-source – Apache stuff is all open-source and I think Hadoop is open-source as far as I know – so ja, absolutely – it's all open-source

RD Thank you for your time and your patience for answering the questions – I appreciate it – is there any questions you got for me?

R1 So you going big data?

Interview Transcript #2

Interviewer name	Romney Davids (RD)	**Interview date**	Thurs, 2017/09/14
Interview start date	17:30	**Interview duration**	23 min 00 sec
Interviewee name	Anonymous (R2)	**Audio filename**	R02.1.20170914
Interviewee occupation	Database Administrator	**Pseudonym**	R2
Notes	Any company names or individual's names have been edited out of the interview transcript.		

RD Recording interview with respondent number 2. Do you give me permission to record this interview?

R2 Yes, sure.

RD Thank you

RSQ1: What is the current skillset and responsibilities of a DBA working in a RDBMS environment?

Q1.1 ***As previously discussed, question number 1 – tell me a bit about yourself. Just give me a brief history of your background as an I.T. professional…***

R2 Okay, I started in … (hmm) …, 1 December 1982 as a computer operator on ISEL mainframes, networks of *[CompanyNameWithheld]*. I operated until 1987 and then became a programmer. Analyst programmer promotion in 1989 and DBA, well different jobs, different companies and then DBA since 2001 at *[CompanyNameWithheld]*. Oh, sorry, no *[CompanyNameWithheld]*.

RD What and where did you study?

R2 Matric in *[SchoolNameWithheld]* High School and Diploma at *[UniversityNameWithheld]*. Started off at *[TechnikonNameWithheld]* and then later on changed to *[UniversityNameWithheld]*

RD You mentioned earlier on that you had a programming background. How did you decide to become a DBA? Is it something you always wanted to do or?

R2 Ja, when you do programming it's the same thing over and over - you have an "if statement" and a "decision" and you execute certain commands. When you do a DBA job, one day is never the same than the previous day. So that was my main reason for going from analytics programming to DBA.

RD Okay

Q1.2 ***You as a DBA, what does your typical day as a DBA look like? How do you start your day? Is there specific routines that you follow or …***

R2 Never the same routine - I'm a Development Application Database Administrator, for a big *[CompanyNameWithheld]* company, working on several different applications and my work is basically given to me by the analysts and programmers or business unit users that requires certain information from the databases. I also have access to the production database for analytical purposes

RD So what is your primary responsibilities as a DBA there?

R2 Data security, data performance of programs, queries in other words

RD And is there certain functions or responsibilities you must perform everyday?

R2 No, not necessarily, just keep up to date with changes … *(could not hear what was said – however there was not much)* … application changes

RD Are there any non-DBA functions that you are required to perform?

R2 Nope, not any

RD And are you required to interface with development, test teams, customers?

R2 Yes, all the time - that's part of my job as well

RD And to what extent?

R2 I assist testers with queries if they need to check data, updates or whatever. Give them queries to do pre-tests and then also for the results after changes were made and then also for the propagation of scripts and so on.

RD With regards to skills of a DBA, what are the core skills that a DBA must have?

R2 Logic, first of all, insight into business practices and the business around the database that you actually working on - very important. And then just general knowledge of the database software itself and then of course performance and so on.

RD And are you able to say what are the current technologies you exposed to or?

R2 Ja, 11g 12c 10g, Unix … is that what you mean?

RD Ja, so whatever technologies you exposed to

R2 Linux, Java - I support java programmers with latest version, VB or dotNet and of course SQL – PL/SQL is development language ja

RD So with regard to core skills and the constant changing of technologies and the emergence of new technologies - how do you as a DBA keep your skills current?

R2 Unfortunately I can't say that at work but at home I use my own PC to you know write small things and learn how to use new technologies - at work we restricted, we not allowed to install any software that's not approved by the company I work for

RD Okay, so you a self-motivated individual

R2 Ja, it's the only way, unfortunately

RD And do you feel it's important for your as a DBA to keep step with the trends

R2 Of course, if you my age you have to be two steps ahead of a normal DBA

Q1.3 ***What other skills or competencies do you think a DBA should have that's not directly related to a DBA?***

R2 Communication skills, first of all, you have to communicate with your users and with the development team or the testers - very important. Again logic, understanding of the business processes

RD And should, do you think a DBA is required to have a development, programming or analytical backgrounds or is it not required?

R2 Both of them - no of course it's both of them if you in a development area. If you do production support which is not application related then you don't need those skills but from my perspective as a development DBA I have to have those skills.

RD Do you think that's a bit different to a production DBA?

R2 Yes, definitely - your production DBA look after hardware and the operating system and the oracle software itself whereas an application DBA you have to do that as a part of your job and the application and business perspective.

RD But non-technical skills like communication, time management, adapting to new technologies etc., those are common across the board

R2 Of course ja

Q1.5 ***With regards to qualifications and training required by a DBA - how important is formal qualifications?***

R2 Okay, I work with a guy that has got 20 years of oracle database experience with no qualifications - except matric and he's doing exceptionally good - he's our application production DBA - he's doing exceptionally good - so I don't think it's necessary to have the qualifications or get certifications from oracle itself but in my case it helped me just by getting confidence that I actually know the structure and how oracle works - did I answer your question?

RD Yes, thank you. So you mentioned certifications, what role do you think vocational training or vendor certifications have within the DBA profession?

R2 Very important role - it's not 100% necessary but if you have that extra training behind you it just helps you along the line. Vendor training I think it's important ... sort of ... 80% important

RD So do you think it helps ... how does it contribute towards the skills of a DBA?

R2 To give you background knowledge that you might need one day and understanding, again the memory structures of the database, you know stuff like that, … (hmm) …, which you won't get from in-house training or talking to other DBAs … they don't generally share that knowledge

RD Alright, great

R2 Is dit reg? *(Afrikaans)*

RD Yes

R2 *(Laughter)*

RD That concludes the first part of the interview – so, the second research sub question relates to the skills and technology requirements of a big data environment

R2 Okay, there I'm not …

RSQ2: What are the skills and technology requirements of a big data environment?

Q2.1 ***Okay, so just in your own words, what are your thoughts on big data, is it something … what are your thoughts on it?***

R2 It's not new technology, first of all, hmm, it …. It's needed because of the volumes of data we having to process and having everything … (hmm) ... available or all the data available as soon as possible … (hmm) … high performance … (hmm) … I don't, because I don't have current experience on big data … (hmm) … I think de-normalizing data if that is part of this ... (hmm) … will make everything faster and more readily available and large volumes but … (hmm) … performance wise when you update data or in batch processes if it's needed then I don't see a place for it in the market … (hmm) … because of the amount of ... (hmm) … programs you have to put behind that ... you know to process data like that …

RD So do you …

R2 Unless it's online data

RD So is it something that you foresee yourself getting involved in?

R2 (Hmm) … not now … not at the place where I'm working now, maybe later you know … If I don't retire before then

RD Is it something that you would like to get involved in?

R2 (Hmm) … I had experience before, like I said in 1987 – 1992 in index-sequential processing which basically is the foundation for big data – except that we didn't have a de-normalized data it was more in third normal form and then processed … (hmm) ... I don't really think that it's the future, maybe I'm wrong, I don't know – for internet type transactions, you know, we need non-batch type data, maybe but not otherwise

RD Okay

R2 Do you think I'm old and out of touch?

RD No, not at all *(Laughter)*

What do you think a typical day of a big data DBA will look like?

R2 I don't know, I've got no idea – maybe very boring … (hmm) … maybe that type of database administrator will also have to be more business type, user, or business person than a technical DBA

RD Okay

R2 The way things are going, ja

Q2.1 ***So, how do you see big data affecting the way DBAs work in the future – would it change the way DBAs work?***

R2 Ja, I think so … (hmm) … the technology is going to be different … (hmm) … if it's a product like oracle if they haven't done it yet, have they? I don't know *(giggles)* … (hmm) … and the background processes I think will take care of the data more than what the old technology relational database admin would have done

RD So so would you think it will have an impact on the responsibilities of DBAs

R2 Yes, it would become less technical, the responsibilities and more business like

RD Okay

R2 So, importance, I don't know

RD And do you think there'll probably be other requirements placed on the DBAs shoulders like security or things like that?

R2 Maybe, security, yes, … (hmm) ... the other thing it's getting more sensitive nowadays, especially in de-normalized form you'll get most probably … (hmm) ... business sensitive data that's available by hitting one table or one set of data … I don't know

RD You wanna take a break?

R2 No. Lets carry on

Q2.3 ***So, what are your thoughts on the skills required in a big data environment? Hmm, do you think …***

R2 Certification obviously

RD Do you think there'll be a requirement to re-skill oneself?

R2 Yes, you will have to most probably … (hmm) ... and then along with that, vendor training or certification courses or something like that

RD Okay, so do you think traditional DBAs already have the skills to manage a big data environment or is it different?

R2 Again, it depends on the experience of these DBAs. A young DBA, 10 years in the job with no other experience as relational databases, they might need to gain more a, what's the word?...

RD skills…

R2 skills, ja, get other skills or learn new skills. Older DBAs that came from a flat file environment or a sequential file environment, they will find it easier to … (hmm) … get the skills to handle big data

RD Okay

R2 That is if I understand big data correctly … *(giggles)*

RD So besides the re-skilling, would there be a need to re-certify?

R2 Yes, I think so. Because of the data security and the integrity … (hmm) ... when you don't have the relational rules prescribing or constraints in this case … (hmm) ... it could be pitfalls in big data … I'm not … *(not sure what was said here but it was not much)*?

RD So, and would you have an idea of what type of certifications it would be or …

R2 (Hmm) … More business like than current technical certifications, OCP, maybe not an OCP more a first level, associate level

RD Okay. What role do you see this vendor certifications have within the big data environment? It might be a …

R2 It should have a big impact … (hmm) … it should be important. Wat vra jy? *(Afrikaans)*

RD What role would this vendor certifications have? Would it be the same as that of a relational environment or …

R2 No, I think more, again more business, security type role than technical

RD And do you think there would be any other skillset required in addition to what a current DBA has? … (hmm) … stuff like coding, analytical, data mining…

R2 Yes of course, ja data mining … (hmm) … again analytical skills I think will be more important than what it is currently for production type DBAs

RD Okay

R2 … because I personally think that production DBAs or like we call them DCO DBAs, that role is going to diminish or become like a clerical work and not technical you know

RD Okay

R2 … depends on the vendor

RD Okay.

Q2.4 ***Big data technologies, … (hmm) … what are your thoughts on the various technologies emerging in this landscape?***

R2 I have no idea

RD Okay

R2 Can't answer that

RD So, do you think these technologies will be short-lived or do you think it will be something that's, that's the way forward

R2 Well, when relational databases took over from network structured type databases we all thought that this is a new thing that's going to be phased out very quickly and look where we ended up with up relational databases. So most probably if the industry accepts it and it's useful and it's safe and the integrity is guaranteed then it might be the future *(lots of background noise and cannot understand what's being said)* - of the volumes *(more background noise drowning out what is being said by R2)*

RD Have you as a DBA started looking at these technologies and competencies required?

R2 (Hmm) … Not data big data as you've got it – I was actually looking at data warehouses – how data warehouses touch data and then I was involved in … (hmm) … like I said *[CompanyNameWithheld]* which I don't know, which I'm not supposed to mention, this company that's doing this type of data transactions, you know, big data, and we spent about a month investigating the viability of using that type of … (hmm) … data, database instead of relational – so that's it – about a month of 8 hour days you know

RD And and is there anything specific that DBAs should do to stay current with the emergence of these new technologies?

R2 Of course we have to study it and … (hmm) … see benefits to tell the companies we work for that you know they should investigate that or go that route. Unfortunately, were I'm now we won't go anywhere … (hmm) … using big data technology

RD Okay

R2 Not, not in the foreseeable future …

RD Okay

RSQ3: What career plans do organisations have for DBAs with regards to big data management?

RD Alright great, we've come to the last research question which looks at career plans for …

R2 I'm *[AgeWithheld]* years old, now you ask me that *(giggles)* … ja …

Q3.1 ***What are the current career plans or paths … (hmm) … that's in place for DBAs in your organization (hmm) …***

R2 Okay, where we are you just a DBA and that's it – there's no career path … (hmm) … but being a specialist job … (hmm) … I think everyone that decides to become a DBA and enjoys doing that job … (hmm) … don't care about a career path – Okay, I've never actually met a DBA that thought about going anywhere else, except on pension of course *(giggles)* … (hmm) … so ja – no career plans

RD So do you think … (hmm) … so there's no career plans or pathing that's been …

R2 Ja, unless they want to … (hmm) … go into a business analytical environment but in my own case I came, I worked as an analyst before I became a DBA … (hmm) … I was a system analyst for *[CompanyNameWithheld]* … (hmm) … oops sorry, and … (hmm) …, so when I started as a DBA that was my career choice – I didn't want to do anything else and it's to this day true - I enjoy my job and … (hmm) … I don't care what I'm called

RD Okay

R2 *(giggles)* it's just salary …

RD So where do you see yourself within the next 5 – 10 years?

R2 … (hmm) …

RD Still working as a DBA?

R2 Still DBA – ja I'm a DBA and I'll die or pension as a DBA

RD Okay

Q3.2 ***… (hmm) … In terms of organizations being able to assist DBAs – what do you think … (hmm) … organizations can do to help DBAs get the needed skills for a big data environment?***

R2 Well, I think it's more the DBAs responsibility to find the information than companies … (hmm) …or what vendors trying to inform people – so first of all companies that employs DBAs or going to use big data … (hmm) … should actually get their DBAs interested in big data technologies and then … (hmm) … sorry …

RD So do you think it's all dependent on the direction of the company or their technology stack?

R2 Yes, of course on the applications, if you have a company that embarks on a new business venture with massive data requirements then obviously they will have to consider that and re-skill their people

RD Okay. Final question, … (hmm) … with the adoption of open source, how do you think the adoption of open-source technologies will assist DBAs in moving towards a role within big data? Do you think it will …

R2 The same role it had when we have programs going open-source … (hmm) … I've actually not seen that – it would be very nice to have open-source technology available on the structure of the databases and how they work, you know … or in this case big data … machines or engines how they actually work in the background … *(maak dit sin?)*

RD Thank you – that concludes the interview … (hmm) … is there any questions you have for me or concerns?

R2 No, nothing

RD Thank you very much

APPENDIX E: Coded interview transcripts for two respondents

(The coded transcripts for all respondents are available upon request)

Interview Transcript #1

Interviewer name:	Romney Davids (RD)	**Interview Date:**	Thurs, 2017-09-14
Interview Start Time:	12:30	**Interview Duration:**	24 min 27 sec
Interviewee Name:	Name Withheld (pseudonym = R1)	**Audio Filename:**	R01.1.20170914
Interviewee Occupation:	Database Administrator Lead		

RD Recording interview with respondent 1. Do you grant me permission to record this interview?
R1 I do.
RD Thank you

RSQ1: What is the current skillset and responsibilities of a DBA working in a RDBMS environment?

	Transcript	Notes
Q1.1	***Tell me a bit about yourself. Just a quick history of your IT background***	
R1	OK. Well my position is Team Lead DBA at *[CompanyNameWithheld]*. I've been a DBA for more than 10 years. Probably about 14 years. I started off as a Oracle Developer. So, I've worked from a development background and I have transitioned to a DBA about 12years ago.	- Leadership position - Extensive DBA experience - Development background - Database developer - Transition from development to DBA role
RD	Okay …	
R1	I've always been involved in Oracle and being a DBA	- Oracle technology expert
RD	And is there something that has made you decide to become a DBA or is it something you always wanted to do?	- Single technology expert? - Development background
R1	(Hmm) …, so I started off as a developer and when I was working as a developer, the DBA guys always seemed to be very knowledgeable and well paid, so I always thought it was a good position to aim for	- Remuneration was incentive to become a DBA - Reason for becoming a DBA - DBAs considered knowledgeable persons
RD	And what did you study to become a developer? Did you specifically study to become a developer or is it just …	- Position aspired towards
R1	No, I initially studied computer science. Um but I did not finish my degree I still had a couple of credits outstanding when I completed so then I ja, I didn't specifically study to become a developer	- Incomplete formal education - Did not study degree with intention of becoming a developer
RD	OK. That's interesting …	- Did he not see a need for a formal education?
Q1.2	***Being a DBA, what does your typical day look like?***	
R1	(Hmm) …	
RD	Is there specific routines you use to start your day or …	
R1	Well ja, usually in the mornings I go through my emails to see if there are any messages from our alert system. We have a system that checks alert logs and makes sure that it does health checks on all of our databases. So I usually go through those and check to make sure that the databases are up and running and if there are any problems then I will resolve those so that's usually the beginning part of my day and the rest of my day is filled with meetings and then attending to other work items, things we have to do for, regarding ja….	- Set morning routine - Automated jobs - Automated health checks - Database uptime important - Meetings & other day-2-day responsibilities
RD	Is there any specific responsibilities that you can define as being primary responsibilities of a DBA?	
R1	Well we primary responsible for providing database services to the floor. I our case that's developers. Developers and testers. So our primary responsibility is to make sure that those databases are up available … (hmm) … and I guess to keep those databases available as much as possible because they all depend on it for their work	- Service RnD department - Developers and testers are main clients - Minimize database downtime - Dependency on services provided by DBAs in order to work
RD	Okay. Are there any non-DBA related functions that you are required to perform? Outside of the general scope of a DBA or that's defined for a DBA?	
R1	(Hmm) …, well as a team lead I guess there's a few management responsibilities that are not DBA related. Then we also in our team we've taken responsibility for some systems that are not DBA related. Specifically, the build system and ja, I	- Management duties - Support non-database related products

Romney Davids R1:1.1:Management Skills
Romney Davids R1:1.1:Extensive Experience
Romney Davids R1:1.1:Development specialist background
Romney Davids R1:1.1:Transition from Developer to DBA Role
Romney Davids R1:1.1:Oracle skills
Romney Davids R1:1.1:Development specialist background
Romney Davids R1:1.1:Professional respect
Romney Davids R1:1.1:Remuneration
Romney Davids R1:1.1:Position aspired towards

Romney Davids R1:1.1:Incomplete tertiary
Romney Davids R1:1.1:Studies not aimed at
Romney Davids R1:1.2:Automated system
Romney Davids R1:1.2:Database uptime
Romney Davids R1:1.2: Daily meetings
Romney Davids R1:1.2:Service related role
Romney Davids R1:1.2:Service R&D department
Romney Davids R1:1.2:Database uptime
Romney Davids R1:1.2:R&D dependency on DBA
Romney Davids R1:1.2:Management skills
Romney Davids R1:1.2:Administor continuous

	suppose specifically the build system I can think of so it's a system that's not related to the DBA field but we have taken on and we support it.	- Support role for continuous integration system
RD	And is there a requirement to interface with other teams in the organization?	
R1	Ja, constantly	- Constant interaction with RnD teams - Job requirement being a service role?
Q1.3	***In terms of core skills for a DBA, are there any specific core skills that you is a requirement for somebody to become a DBA?***	
R1	(Hmm) …	
RD	… or that one should have?	
R1	Well, I guess an overall understanding of IT might be a good start. Then specifically you would need, you need some kind of OS skills, for us most of our work is done in Linux or Unix so those, the OS skills are important. Then kinda getting more specific SQL skills, you need to be able to manipulate databases using SQL, that's how you talk to it. And then there are more specific skills would be database management skills, what's involved in being a DBA, all the admin kinds of stuff that a DBA needs to know.	- General I.T. knowledge - Knowledge of various OS - SQL skills - Linux/Unix based environment - Data manipulation using SQL
RD	Great! Specific technologies you exposed to at the moment?	- Database management skills
R1	In regard to? Just in the database side or…	
RD	Database side or in general as a DBA	
R1	Sjo, I can list a lot, well there's Oracle, Linux, Jenkins, Git, Subversion, Java, Windows	- DBA exposed to many different technologies, not just database
RD	So it's quite specific to a DBA depending on where he works …	
R1	OK, ja, Oracle, mainly Oracle is the main technology	- Oracle technology expert
RD	In terms of your current skills set how do you keep that current? How do you keep abreast of the latest technologies?	
R1	It's all self, what you call self-directed learning. I mean I use it's kind of part of my daily routine is to stay up to date, read the Oracle blogs, read the new features coming out and play around with new features, we are RnD so I do spend quite a	- Self-directed learning - Daily learning - Blogs & new technology features

Romney Davids
R1:1.2:Interaction with other departments

Romney Davids
R1:1.3:General IT knowledge

Romney Davids
R1:1.3: OS skills

Romney Davids
R1:1.3:SQL skills

Romney Davids
R1:1.3:Database management skills

Romney Davids
R1:1.3:Diverse technology exposure

Romney Davids
R1:1.3:Oracle expertise

Romney Davids
R1:1.3:Self-directed learning

Romney Davids
R1:1.3:Online learning

	lot of time when new versions come out installing that and getting to know that. But it's all self-directed	- Continues testing of new software versions & features
RD	So it's a self-taught	
R1	Ja, there are courses and that kind of stuff but that usually comes after the fact, the certification	- Certification courses not primary means of learning; optional extra - Does not seem to consider courses as being very important in order to learn new technologies?
Q1.4	***Are there any other skills or competencies that are not directly related to being a DBA that you should have?***	
R1	You mean non-technical, soft skills …	
RD	Soft skills, non-technical skills or other technical skills	
R1	I think you got to have very good problem solving skills cause lot of the day is solving problems. (Hmm) … you got to be self-motivated and you got to be, I forgotten the word I was looking for …	- Problem-solving skills - Daily requirement to problem-solve - Self-motivated
RD	Communication skills?	- Self-motivated possibly refers to lots idle time during the day?
R1	Ja, you need good communication skills, good team-work skills because you working with a lot of people not just in the DBA team but across the floor. (Hmm) … I think that's all I can think of. Sorry I'm just having a, there one word, if I remember I will let you know …	- Communication skills - Team player - Provide service to many internal customers (RnD)
Q1.5	***Okay … and essential qualifications, is there specific training that one requires to become a DBA?***	
R1	My opinion or I think the industry standard is you would get a Batchelors Degree or some kind of degree tertiary qualification in Computer Science or Computer studies or whatever you call it these days and then that's usually the entry level and then do some kind of Oracle certification. I think is also most jobs require or would prefer Oracle certification	- Tertiary education (degree) a minimum requirement for industry - Vendor certification preferable - Becoming a requirement in the job market?
RD	And in terms of that vocational training or vendor certifications, what role does it play, how does it add to the DBA skills	

Romney Davids
R1:1.3:Self motivation

Romney Davids
R1:1.3:Courses not primary means of learning

Romney Davids
R1:1.4:Problem-solving skills

Romney Davids
R1:1.4:Self motivation

Romney Davids
R1:1.4:Communication skills

Romney Davids
R1:1.4:Interaction with many people

Romney Davids
R1:1.4:Team work

Romney Davids
R1:1.5:Tertiary qualification is minimum job requirement

Romney Davids
R1:1.5:Certification influences job opportunities

R1	Well I guess the main thing is it adds credibility to what the DBA claims, the skills the DBA has, the fact that he has skills, he is certified, it gives him some kind of credibility. I believe that you get more, your skills get more from on the job, day to day work, than going on a one week course	- Certification validates DBA skillset - Adds credibility / proof - Experience better than certification
RD	Okay …	
R1	That's in my experience, but it also sometimes courses are good because they cover gaps that you may have…	- Courses close/cover gaps in knowledge
RD	So it's a forced learning almost …?	
R1	Not forced *(laughter from R1)*, it's on the job, ja, it's forced learning and for any learning to be affective you need to apply it and most people don't apply what they've learned on course	- Application of knowledge is important – i.e. on the job experience - Is theory seldom applied on the job?

RSQ2: What are the skills and technology requirements of a big data environment?

	Transcript	Codes
Q2.1	***The second part of this will deal with the big data environment. What are your thoughts on big data in general. Is it something you see yourself becoming involved in in your career?***	
R1	Well, I don't know if I will specifically become involved in big data but again there's that whole definition of what big data is. Its taken a while, it is becoming clearer what people mean when they say big data. I think in our job we do handle large volumes of data but whether, it doesn't really fit in with what the definition of what big data is, so no I don't see myself going with what is now becoming known as big data unless I move to another company where, that's big data…	- Unsure of future involvement in BD – unclear about definition of BD - Currently handle large DBs as RDBMS DBA (volume of data) - Volume not only criteria for BD - No immediate plans to work with BD
RD	Do you perhaps have any idea of what a big data DBAs day will look like or do you think it will be much the same.	- No exposure to big data in company - Move jobs to be able to work with BD
R1	I would guess fairly similar although the one difference – one of the differences we development DBAs whereas big data DBAs I imagine are going to be more	- BD DBA similar role to RDBMS DBA - BD DBA will not do application DBA work
	production focus because big data is about collecting data that's coming in all the time so I guess big data dba will be more production focused and will be having to monitor data feeds, really constantly, because big data, your data feeds don't stop, they just come at you all the time, that's how I understand it, so there'll be more, there'll be less development and more production kind of stuff, monitoring, making sure systems are running that kind of thing – if that makes sense	- BD DBA has production focus - BD is constant stream of data - Monitor data feeds - 24 hr a day data streams - BD is more production type environment - Focus on DB uptime
Q2.2	***How do you see big data affecting the way DBAs work in the future? I think you probably touched on that a bit in your previous answer.***	
R1	Ja, it's a hard question to answer	- Unsure how BD DBA will be different to RDBMS DBA
RD	Ok, lets put it this way …	
R1	Look I think any DBA going forward, anyone who stays in the DBA field needs to be up to speed with what big data is and maybe big data is going to become less of a side-show and normal database work and big data will become, will kind of merge together. Already you seeing with Oracle they got all these big data connectors and that kind of thing. So big data is just becoming part of normal day-2-day database work. Maybe.	- All DBAs must stay current with BD - Unsure whether BD will become mainstream or RDBMS to continue to be main stream - Possible merger between RDBMS & BD - Vendors more focused on BD technologies - BD more common place
RD	And would that affect, that impact on DBAs responsibilities. Would it be any different to what DBAs do in a relational environment?	
R1	I don't think so, just again, going back to the incoming data feeds, if you working, it just depends if you in a dev or production kind of environment. I think one of the big things if you become a big data DBA will be taking care of data feeds, data coming in. Because obviously you got to process, big data you got a lot of data to process, so your data feeds can't back up, if you know what I mean – a big hose pipe, it just keeps coming	- No difference in RDBMS & BD DBA responsibilities - Responsibilities dependent on environment (app/dev vs prod) - DB uptime is critical part of BD environment – constant incoming data
RD	And in terms of specific requirements like security and things like that do you there'll be more focus on that or	
R1	I think ja, I think definitely, they are becoming more important in the industry is security and privacy as well. So, but I don't know how that, where privacy fits in a DBAs job because we don't really manage the content of data, we just manage the data from a physical point of view	- Security important in BD - Privacy important in BD - DBA focus is not on data content itself but management of data - Manage data from physical point of view
Q2.3	***And your thoughts on the skills of a big data DBA? Would you foresee a requirement to reskill?***	
R1	You'd need to expand your skillset, obviously the big data tools are different to what we currently use here – so you'd need to learn things like Hadoop for example or whatever the big data – so there will be a reskilling involved but I don't think, it's not a case of switching skillsets, it's a case of expanding skillset – there's a lot of knowledge which we've learned as DBAs that is very applicable – that's why I say I don't see the day-2-day job being that different just a different set of tools and bigger datasets	- Expand current skillset - BD tools different to RDBMS tools - Learn new toolsets for BD (Hadoop) - Current skillset still relevant/required – just add additional skills for BD - RDBMS DBA knowledge important - RDBMS & BD DBA daily job/focus the same – only different tools
RD	Do you think the traditional DBA or the RDBMS DBA have sufficient skills to manage this big data environments?	
R1	With the right tools I think that they do, ja	- RDBMS DBA skillset sufficient for BD environment – learn new tools
RD	And would there be a possibility or requirement to re-certify themselves?	
R1	Probably advisable, ja – because, if you going, as I say, it gives you the additional credibility when you're out there in the market – so if you do reskill then it make sense to re-certify yourself – otherwise you wasting your money really	- BD certification advisable - Certifications add credibility to skills in market - Certification linked to skillset
RD	And would there be any additional skillsets or other skillset required that is currently expected – or is there any other skillset that a big data DBA will be required to have? That's any different to …	- Cost factor *(probably referring more to getting the additional skills???)*
R1	… to current ones – You know apart from different tools, I don't think so unless you starting to look at actual data science and data mining and all that kind of stuff – but I'm excluding those things from this – I'm thinking you get more data engineers and you get data – hmm, and that the way it seems to be moving you	- RDBMS & BD DBA skillset similar - RDBMS & BD DBA use different tools - Data science / mining requires different skillset to BD

Romney Davids R1:1.5:Certification validates
Romney Davids R1:1.5:Prefers experience over
Romney Davids R1:1.5:Vendor training covers
Romney Davids R1:1.5:On-the-job experience
Romney Davids R1:2.1:Unsure about future
Romney Davids R1:2.1:Administor large volume
Romney Davids R1:2.1:Volume not sole criteria
Romney Davids R1:2.1:No big data
Romney Davids R1:2.1:Change jobs to get big
Romney Davids R1:2.1:Very similar role to
Romney Davids R1:2.1:Big data DBA has
Romney Davids R1:2.1:Continuous flow of data
Romney Davids R1:2.1:Monitoring of data feeds
Romney Davids R1:2.1:24x7 data streams
Romney Davids R1:2.1:Production oriented
Romney Davids R1:2.2:DBA must stay current
Romney Davids R1:2.2:Potential merger
Romney Davids R1:2.2:Big data more common
Romney Davids R1:2.2:Same responsibilities
Romney Davids R1:2.2:Environment determines
Romney Davids R1:2.2:Data feeds
Romney Davids R1:2.2:Constant data coming in
Romney Davids R1:2.2:Database uptime
Romney Davids R1:2.2:Security
Romney Davids R1:2.2:Privacy important in big
Romney Davids R1:2.2:Manage physical aspect
Romney Davids R1:2.3:Expand current skillset
Romney Davids R1:2.3:Learn new toolsets
Romney Davids R1:2.3:RDBMS DBA skillset still
Romney Davids R1:2.3:Only difference is tools
Romney Davids R1:2.3:DBA has skills to
Romney Davids R1:2.3:Learn new tools
Romney Davids R1:2.3:Certification adds market
Romney Davids R1:2.3:Big data certification
Romney Davids R1:2.3:Re-certify if re-skilling
Romney Davids R1:2.3:Data science/mining
Romney Davids R1:2.3:Similar skillset as RDBMS
Romney Davids R1:2.3:DBA to become Data

	get data engineers and data scientists – and the data engineers are us, going to be the DBAs – so in terms of skillsets, probably not, I think it's the same kind of job, just different tools and …. I don't think there's a big – as I say the tools are different	- DBA to become data engineer in BD environment - RDBMS DBA skillset sufficient for BD - Different tools for BD but skillset still applicable
Q2.4	***Your thoughts on the various technologies within the big data landscape – are they, do you think they'll be short-lived or is it the way forward for DBAs?***	
R1	I wouldn't, well I haven't and I wouldn't to rush off to become a big data, to drop everything and learn all the big data toolsets – but definitely there are some toolsets that are standing out from the pack like Hadoop and Apache Spark and R I think is one of the other big ones – so there are definitely now becoming – it's becoming clear what the tools are that are going to move forward – so if I was moving that way then definitely there are things you can follow up on – that answer your question	- Not rushed to learn BD tools - RDBMS toolset still relevant - Some tools more prominent/dominant than others
RD	Have you started looking those technologies or competencies required for a big data?	
R1	I have, I've kind of messed around with Hadoop and but ja, to some degree, just so that I'm familiar what – how it all hangs together – I haven't done any practical kind of work with it no	- No real-world experience on BD tools - Experimented with Hadoop – inquisitive
RD	And with all these new technologies emerging, **how should DBAs stay current** with all the emergence of all these technologies? Would it be much of the same in terms of the way they currently upskilling themselves or keep themselves current or …	
R1	I think we still at a stage where DBAs need to decide if they want to become big data DBAs or whether they want to just be RDBMS DBAs – I think there's still a distinction but later on it will merge, so at this point if you decide you want to be a DBA, a big data DBA, then you need to go an move to those skillsets and I think, but if you are currently wanting to stay as a RDBMS DBA I don't think you need to investigate that stuff too much more than knowing what it's about – well that's kind of the approach I've taken	- DBA to decide on development path (RDBMS or BD) - Possible merger between RDBMS & BD - BD DBA to focus on new toolsets *(here skillset is mentioned but earlier mentioned it's only tools that different?)* - RDBMS DBA only need general understanding of what BD toolsets about

RSQ3: What career plans do organisations have for DBAs with regards to big data management?

	Transcript	Notes
Q3.1	***The last section will deal with career plans and the way forward for DBAs. What are the current career development plans or paths that's in place within your organization for DBAs? Would you be able to pursue a career in big data?***	
R1	Not in our current organization – I don't think there's any scope for big data DBAs and there are no career plans	- No BD career path/plans for RDBMS DBA in current organization
RD	And in terms of career pathing has anything been … okay you've …	
R1	no, but I can cover it again, so definitely I think at this stage you got to decide whether and I'm not talking about in our company, I think in the industry in general, If you want to go with the big data DBA you need to make that choice and then you need to follow that path - I think the path is independent, not independent but it does split from being a relational database DBA although you could do both	- RDBMS DBA to decide on career path – BD vs RDBMS DBA path - Seems to indicate that BD & RDBMS DBA are two separate paths
RD	How would you describe your prospects of moving towards a role within a big data environment?	- Possibility of doing RDBMS & BD DBA work??
R1	If I was going to go that route I think the prospects would be good but it would be a case of specifically going and investigating and reskilling – there would be a reskilling and re-certification involved but I think I would quite easily make the transition – I'm just not familiar with the industry because as of yet I don't think big data industry, and this can be completely wrong, in cape town has grown much	- RDBMS DBA prospects good if wanting to move to BD environment - Should not be a difficult transition for RDBMS DBA to move to BD
RD	Just going back to that question, do you think it's your experience as a traditional DBA that would stand you in good stead to become a big data DBA or is it …	- Re-skilling & re-certification requirement for RDBMS DBA moving to BD environment
R1	No, definitely - as I say I don't think there's – I think just to kind of sum up what I think where lot of the questions we're answering up – I think someone who's been a traditional DBA could make the move quite easily but they would have to reskill in addition to their current skillset – but anyone who was kind of trying to move forward and trying to decide which – there is at some point, I guess you got to choose whether you – the two paths are split – you know what I'm saying – currently at some point you either remain a traditional DBA with your current skills and you just keep that current or if you decide to become a big data DBA, your prospects are good but you have to take, you have to reskill and re-certify- ja, it's not going to be difficult it's just something that has to be done	- BD industry now evolved in CPT – hence prospects of moving to BD environment not very good at present - RDBMS DBA experience & knowledge key to moving to BD environment - RDBMS DBA only needs to enhance current skillset - RDBMS & BD DBA are 2 separate paths - RDBMS DBA to reskill & re-certify – enhance current skillset
RD	And where would you see yourself in 5-10 years' time	
R1	Retired, no I would probably remain traditional at this point just basically because this company is not going big data in the foreseeable future – that being said, we'll see what the industry is doing as well – at this point, big data is still a bit of a, I mean it's not a buzzword it is becoming real but I don't where the jobs are, you know	- No future BD prospects in current company – hence remain RDBMS DBA - BD still a growing industry – seems to indicate that there's no jobs in CPT yet - Possibly implies that he has been looking?
Q3.2	***And is there anything specific organizations can do to help DBAs to get the required skills***	
R1	Well it just depends of the organization is going that way but there must be certain training, certifications for things like Cassandra, for Hadoop or Apache Spark – so if a company is, if this company were to tomorrow have a sudden big data requirement then there are definitely certifications and trainings that they can	- DBA prospects for BD is dependent on company direction

Romney Davids R1:2.3:No difference in RDBMS

Romney Davids R1:2.4:No rush to learn big data
Romney Davids R1:2.4:RDBMS will continue to
Romney Davids R1:2.4:Some big data tools more

Romney Davids R1:2.4:No real-world big data
Romney Davids R1:2.4:Limited exposure with

Romney Davids R1:2.4:DBA to decide career
Romney Davids R1:2.4:Possible merger between
Romney Davids R1:2.4:Keep updated with

Romney Davids R1:3.1:No big data career plans in company

Romney Davids R1:3.1:Big data DBA separate path to RDBMS

Romney Davids R1:3.1:Decide which path to pursue

Romney Davids R1:3.1:Can do both RDBMS & big data DBA work

Romney Davids R1:3.1:Has ability to transition to big data

Romney Davids R1:3.1:Re-skilling & re-certification required

Romney Davids R1:3.1:Industry popularity/size makes job prospects low

Romney Davids R1:3.1:RDBMS DBA
Romney Davids R1:3.1:Only needs to enhance

Romney Davids R1:3.1:RDBMS & big data are
Romney Davids R1:3.1:Re-skilling & re-
Romney Davids R1:3.1:Expand current skillset

Romney Davids R1:3.1:Remain RDBMS because
Romney Davids R1:3.1:Lack of advertised jobs

Romney Davids R1:3.2:No current big data
Romney Davids R1:3.2:Upskill through courses

	send people on out there I would imagine because that is now fairly normal thing big data	- Company can send DBA on re-skilling courses and certifications to assist transition to BD
RD	Okay. And lastly the adoption of open-source technologies within the organization, would that assist DBAs in moving that direction or is it still dependent on …	
R1	Ja, I think a lot of the, the majority actually if I think of it of the big data stuff is open-source – Apache stuff is all open-source and I think Hadoop is open-source as far as I know – so ja, absolutely – it's all open-source	- BD technologies primarily open-source - Agrees that adoption of open-source can assist RDBMS towards a transition to BD
RD	Thank you for your time and your patience for answering the questions – I appreciate it – is there any questions you got for me?	- Company adoption of open-source technologies can assist DBAs in transition???
R1	So you going big data?	

Romney Davids
R1:3.2:Open-source adoption can aid DBA

Interview Transcript #2

Interviewer name:	Romney Davids (RD)	**Interview Date:**	2017-09-14
Interview Start Time:	17:30	**Interview Duration:**	23 min 00 sec
Interviewee Name:	Name Withheld (pseudonym = R2)	**Audio Filename:**	R02.1.20170914
Interviewee Occupation:	Database Administrator		

RD Recording interview with respondent number 2. Do you give me permission to record this interview?
R2 Yes, sure.
RD Thank you

RSQ1: What is the current skillset and responsibilities of a DBA working in a RDBMS environment?

	Transcript	Notes
Q1.1	***As previously discussed, question number 1 – tell me a bit about yourself. Just give me a brief history of your background as an I.T. professional…***	- Entire career in I.T.
R2	Okay, I started in … (hmm) …, 1 December 1982 as a computer operator on ISEL mainframes, networks of *[CompanyNameWithheld]*. I operated until 1987 and then became a programmer. Analyst programmer promotion in 1989 and DBA, well different jobs, different companies and then DBA since 2001 at *[CompanyNameWithheld]*. Oh, sorry, no *[CompanyNameWithheld]*.	- Mainframe operator background - Development background – transition from development to DBA role - Experienced DBA
RD	What and where did you study?	
R2	Matric in *[SchoolNameWithheld]* High School and Diploma at *[UniversityNameWithheld]*. Started off at *[TechnikonNameWithheld]* and then later on changed to *[UniversityNameWithheld]*	- Formal tertiary education at various institutions
RD	You mentioned earlier on that you had a programming background. How did you decide to become a DBA? Is it something you always wanted to do or?	

Romney Davids
R2:1.1:Mainframe operator background

Romney Davids
R2:1.1:Development specialist background

Romney Davids
R2:1.1:Development to DBA transition

Romney Davids
R2:1.1:Senior DBA experience

Romney Davids
R2:1.1:Tertiary qualification (National Diploma)

R2	Ja, when you do programming it's the same thing over and over - you have an "if statement" and a "decision" and you execute certain commands. When you do a DBA job, one day is never the same than the previous day. So that was my main reason for going from analytics programming to DBA.	- Development became repetitive routine – bored - DBA provided a more challenging environment
RD	Okay	
Q1.2	***You as a DBA, what does your typical day as a DBA look like? How do you start your day? Is there specific routines that you follow or …***	- Application DBA role
R2	Never the same routine - I'm a Development Application Database Administrator, for a big *[CompanyNameWithheld]* company, working on several different applications and my work is basically given to me by the analysts and programmers or business unit users that requires certain information from the databases. I also have access to the production database for analytical purposes	- No set routine – constantly changing environment - Prescribed set of work from business and development teams - Analytical role also part of job function on production database
RD	So what is your primary responsibilities as a DBA there?	
R2	Data security, data performance of programs, queries in other words	- Data security, performance tuning, assist development teams with SQL queries
RD	And is there certain functions or responsibilities you must perform everyday?	
R2	No, not necessarily, just keep up to date with changes … *(could not hear what was said – however there was not much)* … application changes	- Constantly changing environment - Work in conjunction with changes made by development team
RD	Are there any non-DBA functions that you are required to perform?	
R2	Nope, not any	- No non-DBA related responsibilities
RD	And are you required to interface with development, test teams, customers?	
R2	Yes, all the time - that's part of my job as well	- Constant interaction with end-users
RD	And to what extent?	
R2	I assist testers with queries if they need to check data, updates or whatever. Give them queries to do pre-tests and then also for the results after changes were made and then also for the propagation of scripts and so on.	- Close working relationship with development and testing teams - Provide guidance role by writing queries for testers
Q1.3	***With regards to skills of a DBA, what are the core skills that a DBA must have?***	

Romney Davids R2:1.1:Challenging environment

Romney Davids R2:1.2:No set routine

Romney Davids R2:1.2:Interface with Analysts

Romney Davids R2:1.2:Interface with

Romney Davids R2:1.2:Data security skills

Romney Davids R2:1.2:Performance tuning

Romney Davids R2:1.2:Collaborate with

Romney Davids R2:1.2:No non-dba functions

Romney Davids R2:1.2:Interaction with other

Romney Davids R2:1.2:SQL skills

R2	Logic, first of all, insight into business practices and the business around the database that you actually working on - very important. And then just general knowledge of the database software itself and then of course performance and so on.	- Business knowledge - Logical thinker - General understanding of the database technology
RD	And are you able to say what are the current technologies you exposed to or?	- Performance tuning
R2	Ja, 11g 12c 10g, Unix … is that what you mean?	- Extensive exposure to many versions of the database technology
RD	Ja, so whatever technologies you exposed to	
R2	Linux, Java - I support java programmers with latest version, VB or dotNet and of course SQL – PL/SQL is development language ja	- Unix / Linux / OS Skills - Programming experience in Java, VB, dotNet to assist development teams (development knowledge requirement)
RD	So with regard to core skills and the constant changing of technologies and the emergence of new technologies - how do you as a DBA keep your skills current?	
R2	Unfortunately I can't say that at work but at home I use my own PC to you know write small things and learn how to use new technologies - at work we restricted, we not allowed to install any software that's not approved by the company I work for	- Must have knowledge of database query language - Strict work environment – not allowed to explore or self-learn new technologies
RD	Okay, so you a self-motivated individual	- Self-directed learning in private capacity
R2	Ja, it's the only way, unfortunately	- Prescribed list of software allowed on workstations
RD	And do you feel it's important for your as a DBA to keep step with the trends	
R2	Of course, if you my age you have to be two steps ahead of a normal DBA	- Self-directed learning is only means of improving skill-set
Q1.4	***What other skills or competencies do you think a DBA should have that's not directly related to a DBA?***	- More difficult to grasp technology/database concepts with age
R2	Communication skills, first of all, you have to communicate with your users and with the development team or the testers - very important. Again logic, understanding of the business processes	- Considers age as a factor to grasping new concepts - Communication skills important (interaction with end-users)
RD	And should, do you think a DBA is required to have a development, programming or analytical backgrounds or is it not required?	- Logic - Understand the business processes

Romney Davids R2:1 3:Logical thinker
Romney Davids R2:1 3:Business knowledge
Romney Davids R2:1 3:Database knowledge
Romney Davids R2:1 3:Performance tuning
Romney Davids R2:1 3:Diverse technology
Romney Davids R2:1 3:Upskill in private capacity
Romney Davids R2:1 3:Restricted working
Romney Davids R2:1 3:Self motivation
Romney Davids R2:1 3:Age influences learning
Romney Davids R2:1 4:Communication skills
Romney Davids R2:1 4:Logical thinker
Romney Davids R2:1 4:Business knowledge

R2	Both of them - no of course it's both of them if you in a development area. If you do production support which is not application related then you don't need those skills but from my perspective as a development DBA I have to have those skills.	- Is logic an indicator for problem-solving skills? - Development/analytical skills required for application DBA environment
RD	Do you think that's a bit different to a production DBA?	- Production DBA environment don't need development/analytical skills
R2	Yes, definitely - your production DBA look after hardware and the operating system and the oracle software itself whereas an application DBA you have to do that as a part of your job and the application and business perspective	- Production DBA works with hardware, OS and database software
RD	But non-technical skills like communication, time management, adapting to new technologies etc., those are common across the board	- Application DBA incorporates production DBA tasks plus development & business processes
R2	Of course ja	- Broad range of soft-skills a requirement
Q1.5	***With regards to qualifications and training required by a DBA - how important is formal qualifications?***	
R2	Okay, I work with a guy that has got 20 years of oracle database experience with no qualifications - except matric and he's doing exceptionally good - he's our application production DBA - he's doing exceptionally good - so I don't think it's necessary to have the qualifications or get certifications from oracle itself but in my case it helped me just by getting confidence that I actually know the structure and how oracle works - did I answer your question?	- No formal or vocational training does not prevent one from performing as a DBA - On-the-job experience is imperative - Certification boosts confidence and validate knowledge of database technology
RD	Yes, thank you. So you mentioned certifications, what role do you think vocational training or vendor certifications have within the DBA profession?	
R2	Very important role - it's not 100% necessary but if you have that extra training behind you it just helps you along the line. Vendor training I think it's important … sort of … 80% important	- Bit of a contradiction to the previous statement – now mentioning that vendor certifications play an important role
RD	So do you think it helps … how does it contribute towards the skills of a DBA?	- Not a requirement to do the work as a DBA but still important
R2	To give you background knowledge that you might need one day and understanding, again the memory structures of the database, you know stuff like	- Seems though that it's an optional extra but plays an important role

Romney Davids R2:1 4:Analytical skills
Romney Davids R2:1 4:Programming skills
Romney Davids R2:1 4:Skillset dependent on
Romney Davids R2:1 5:Tertiary qualification not
Romney Davids R2:1 5:Certifications or
Romney Davids R2:1 5:Certification provides
Romney Davids R2:1 5:Positive attitude towards
Romney Davids R2:1 5:Vendor training
Romney Davids R2:1 5:Vendor training covers

	that, … (hmm) …, which you won't get from in-house training or talking to other DBAs … they don't generally share that knowledge	- Vocational training covers gaps in one's knowledge / background knowledge
RD	Alright, great	- Certain types of knowledge withheld by DBAs – don't share easily?
R2	Is dit reg? *(Afrikaans)*	
RD	Yes	- Do DBAs not share because the courses are so expensive?
R2	*(Laughter)*	
RD	That concludes the first part of the interview – so, the second research sub question relates to the skills and technology requirements of a big data environment	
R2	Okay, there I'm not …	

Romney Davids R2:1 5:Lack of knowledge sharing

RSQ2: What are the skills and technology requirements of a big data environment?

Transcript		Notes
Q2.1	***Okay, so just in your own words, what are your thoughts on big data, is it something … what are your thoughts on it?***	
R2	It's not new technology, first of all, hmm, it …. It's needed because of the volumes of data we having to process and having everything … (hmm) … available or all the data available as soon as possible … (hmm) … high performance … (hmm) … I don't, because I don't have current experience on big data … (hmm) … I think de-normalizing data if that is part of this … (hmm) … will make everything faster and more readily available and large volumes but … (hmm) … performance wise when you update data or in batch processes if it's needed then I don't see a place for it in the market … (hmm) … because of the amount of … (hmm) … programs you have to put behind that … you know to process data like that …	- BD not new - BD related to volume - Data availability has become important - Performance is important - No BD experience - Not working with BD at present - Change in database structure - Suggests de-normalizing / flattening the DB structure - BD data environment could become complex
RD	So do you …	

Romney Davids R2:2 1:Big data not new
Romney Davids R1:2 1:Big data linked with volume
Romney Davids R2:2 1:No big data exposure
Romney Davids R2:2 1:Database structure should be de-normalized
Romney Davids R2:2 1:Complex environment

R2	Unless it's online data	- Too much backend systems to manage the data in a BD environment
RD	So is it something that you foresee yourself getting involved in?	
R2	(Hmm) … not now … not at the place where I'm working now, maybe later you know … If I don't retire before then	- No exposure/experience in BD in current organization
RD	Is it something that you would like to get involved in?	- Does possibly see himself/herself working with BD in the future
R2	(Hmm) … I had experience before, like I said in 1987 – 1992 in index-sequential processing which basically is the foundation for big data – except that we didn't have a de-normalized data it was more in third normal form and then processed … (hmm) … I don't really think that it's the future, maybe I'm wrong, I don't know – for internet type transactions, you know, we need non-batch type data, maybe but not otherwise	- Exposure to BD dependent on company/environment - BD not new / considers initial DB experience still relevant to BD today - Index-sequential files foundation of BD
RD	Okay	- BD nothing new / not the future
R2	Do you think I'm old and out of touch?	- Questions his own knowledge/exposure of current trends
RD	No, not at all *(Laughter)* What do you think a typical day of a big data DBA will look like?	- Considers his age as a factor for not keeping up-to-date with current technology trends
R2	I don't know, I've got no idea – maybe very boring … (hmm) … maybe that type of database administrator will also have to be more business type, user, or business person than a technical DBA	- No insight of BD environment - BD DBA will focus more on business requirements/processes than technical
RD	Okay	
R2	The way things are going, ja	- Boring environment - Non-technical type DBA (business / process)
Q2.2	***So, how do you see big data affecting the way DBAs work in the future – would it change the way DBAs work?***	
R2	Ja, I think so … (hmm) … the technology is going to be different … (hmm) … if it's a product like oracle if they haven't done it yet, have they? I don't know *(giggles)* … (hmm) … and the background processes I think will take care of the data more than what the old technology relational database admin would have done	- BD requires different technology stack - BD technology stack affects way DBA work - BD technology more autonomous than RDBMS
RD	So so would you think it will have an impact on the responsibilities of DBAs	

R2	Yes, it would become less technical, the responsibilities and more business like	- Less technical DBA required for BD
RD	Okay	- BD DBA will have a more business focus
R2	So, importance, I don't know	
RD	And do you think there'll probably be other requirements placed on the DBAs shoulders like security or things like that?	
R2	Maybe, security, yes, … (hmm) ... the other thing it's getting more sensitive nowadays, especially in de-normalized form you'll get most probably … (hmm) … business sensitive data that's available by hitting one table or one set of data … I don't know	- Security important in BD environment - Sensitive data - Sensitive data easily accessible because of de-normalized state
RD	You wanna take a break?	
R2	No. Lets carry on	
Q2.3	***So, what are your thoughts on the skills required in a big data environment? Hmm, do you think …***	- Vendor certification important
R2	Certification obviously	
RD	Do you think there'll be a requirement to re-skill oneself?	- Reskilling requirement for BD environment
R2	Yes, you will have to most probably … (hmm) ... and then along with that, vendor training or certification courses or something like that	- Vendor courses/certification for BD environment
RD	Okay, so do you think traditional DBAs already have the skills to manage a big data environment or is it different?	
R2	Again, it depends on the experience of these DBAs. A young DBA, 10 years in the job with no other experience as relational databases, they might need to gain more a, what's the word?...	- Experience important factor to transition
RD	skills…	
R2	skills, ja, get other skills or learn new skills. Older DBAs that came from a flat file environment or a sequential file environment, they will find it easier to … (hmm) … get the skills to handle big data	- Transition to BD environment easier for experienced DBAs - DBAs with flat / sequential file experience will find transition easier to BD environment
RD	Okay	

R2	That is if I understand big data correctly … *(giggles)*	- Questions own understanding of what is BD
RD	So besides the re-skilling, would there be a need to re-certify?	
R2	Yes, I think so. Because of the data security and the integrity … (hmm) ... when you don't have the relational rules prescribing or constraints in this case … (hmm) ... it could be pitfalls in big data … I'm not … *(some words inaudible)*?	- Re-certification required - Less rigid architecture than RDBMS - RDBMS provides better protection for data – constraints etc. provides some form of security
RD	So, and would you have an idea of what type of certifications it would be or …	
R2	(Hmm) … More business like than current technical certifications, OCP, maybe not an OCP more a first level, associate level	- BD certification less technical / more focused on business
RD	Okay. What role do you see this vendor certifications have within the big data environment? It might be a …	- Not sure of what certifications will be
R2	It should have a big impact … (hmm) … it should be important. Wat vra jy? *(Afrikaans)*	- Vendor certification critical in BD environment
RD	What role would this vendor certifications have? Would it be the same as that of a relational environment or …	
R2	No, I think more, again more business, security type role than technical	- Vendor certifications more focused on business requirements / less technical
RD	And do you think there would be any other skillset required in addition to what a current DBA has? … (hmm) … stuff like coding, analytical, data mining…	
R2	Yes of course, ja data mining … (hmm) … again analytical skills I think will be more important than what it is currently for production type DBAs	- BD DBA must have data mining skills - BD DBA must have analytical skills
RD	Okay	
R2	… because I personally think that production DBAs or like we call them DCO DBAs, that role is going to diminish or become like a clerical work and not technical you know	- Production DBA role to become less prevalent - Less technical requirement on DBA
RD	Okay	- DBA skillset vendor specific
R2	… depends on the vendor	
RD	Okay.	

Romney Davids R2:2.1:No big data
Romney Davids R2:2.1:No big data experience
Romney Davids R2:2.1:Exposure to index-
Romney Davids R2:2.1:Index-sequential files is
Romney Davids R2:2.1:Big data won't last
Romney Davids R2:2.1:No insight into big data
Romney Davids R2:2.1:Boring environment
Romney Davids R2:2.1:DBA more business
Romney Davids R2:2.2:Different technology in
Romney Davids R2:2.2:Work differently becau
Romney Davids R2:2.2:Autonomous databases

Romney Davids R2:2.2:Less technical requirement
Romney Davids R2:2.2:More business focus for DBA
Romney Davids R2:2.2:Sensitive data
Romney Davids R2:2.2:Security
Romney Davids R2:2.3:Certification very important
Romney Davids R2:2.3:Re-skill in big data
Romney Davids R2:2.3:Big data certification advisable
Romney Davids R2:2.3:Transition influenced by experience
Romney Davids R2:2.3:Transition easier for older DBAs

Romney Davids R2:2.3:Questions understandi
Romney Davids R2:2.3:Less rigid architecture
Romney Davids R2:2.3:Re-certification required
Romney Davids R2:2.3:Less secure environme
Romney Davids R2:2.3:Certification more
Romney Davids R2:2.3:Less technical
Romney Davids R2:2.3:Certification critical in BD
Romney Davids R2:2.3:Greater business and
Romney Davids R2:2.3:Upskill in data analysis
Romney Davids R2:2.3:Upskill in data mining
Romney Davids R2:2.3:Production DBA role to
Romney Davids R2:2.3:Less technical
Romney Davids R2:2.3:Skillset will be vendor

Q2.4	***Big data technologies, … (hmm) … what are your thoughts on the various technologies emerging in this landscape?***	- Not informed of BD technologies/trends
R2	I have no idea	- Lack of interest in BD technologies
RD	Okay	
R2	Can't answer that	
RD	So, do you think these technologies will be short-lived or do you think it will be something that's, that's the way forward	
R2	Well, when relational databases took over from network structured type databases we all thought that this is a new thing that's going to be phased out very quickly and look where we ended up with up relational databases. So most probably if the industry accepts it and it's useful and it's safe and the integrity is guaranteed then it might be the future *(some words inaudible)* - of the volumes *(some words inaudible)*	- BD technology growth dependent on industry acceptance - Security is important
RD	Have you as a DBA started looking at these technologies and competencies required?	
R2	(Hmm) … Not data big data as you've got it – I was actually looking at data warehouses – how data warehouses touch data and then I was involved in … (hmm) … like I said *[CompanyNameWithheld]* which I don't know, which I'm not supposed to mention, this company that's doing this type of data transactions, you know, big data, and we spent about a month investigating the viability of using that type of … (hmm) … data, database instead of relational – so that's it – about a month of 8 hour days you know	- No interest in BD - Interested in RDBMS / data warehousing - Have not done any investigation into BD technologies yet - Exposed to BD project in company to investigate viability of BD technologies - Seems like project was a dead-end
RD	And and is there anything specific that DBAs should do to stay current with the emergence of these new technologies?	
R2	Of course we have to study it and … (hmm) … see benefits to tell the companies we work for that you know they should investigate that or go that route. Unfortunately, were I'm now we won't go anywhere … (hmm) … using big data technology	- Self-directed learning for DBA / stay current with emerging technologies / DBA plays advisor role to company - No BD prospects in current company

Romney Davids R2:2.4:No knowledge of big data
Romney Davids R2:2.4:Little interest in big data
Romney Davids R2:2.4:Big data technology
Romney Davids R2:2.4:Big data technology
Romney Davids R2:2.4:Interest in RDBMS/data
Romney Davids R2:2.4:Limited exposure with
Romney Davids R2:2.4:Keep updated with
Romney Davids R2:2.4:DBA has advisor role
Romney Davids R2:2.4:No big data

RD	Okay	
R2	Not, not in the foreseeable future …	- No future BD plans
RD	Okay	

RSQ3: What career plans do organisations have for DBAs with regards to big data management?

	Transcript	Notes
RD	Alright great, we've come to the last research question which looks at career plans for …	
R2	I'm *[AgeWithheld]* years old, now you ask me that *(giggles)* … ja …	
Q3.1	***What are the current career plans or paths … (hmm) … that's in place for DBAs in your organization (hmm) …***	
R2	Okay, where we are you just a DBA and that's it – there's no career path … (hmm) … but being a specialist job … (hmm) … I think everyone that decides to become a DBA and enjoys doing that job … (hmm) … don't care about a career path – Okay, I've never actually met a DBA that thought about going anywhere else, except on pension of course *(giggles)* … (hmm) … so ja – no career plans	- No career pathing / growth for DBA - DBA a specialist job – DBAs don't intend to change career paths once settled - DBA for life until retirement
RD	So do you think … (hmm) … so there's no career plans or pathing that's been …	
R2	Ja, unless they want to … (hmm) … go into a business analytical environment but in my own case I came, I worked as an analyst before I became a DBA … (hmm) … I was a system analyst for *[CompanyNameWithheld]* … (hmm) … oops sorry, and … (hmm) …, so when I started as a DBA that was my career choice – I didn't want to do anything else and it's to this day true - I enjoy my job and … (hmm) … I don't care what I'm called	- Potential career path for DBA to move into is business analyst - Analytical role / background prior to becoming a DBA - DBA career by choice - Enjoys job as DBA / job satisfaction
RD	Okay	
R2	*(giggles)* it's just salary …	- Remuneration is important consideration

Romney Davids R2:3.1:No career pathing towards big data
Romney Davids R2:3.1:Enjoyment for job
Romney Davids R2:3.1:DBAs don't change career paths once settled
Romney Davids R2:3.1:Business Analyst potential career path for DBA
Romney Davids R2:3.1:Analyst background
Romney Davids R2:3.1:DBA by choice
Romney Davids R2:3.1:Enjoyment for job
Romney Davids R2:3.1:Remuneration is important

RD	So where do you see yourself within the next 5 – 10 years?	
R2	… (hmm) …	
RD	Still working as a DBA?	
R2	Still DBA – ja I'm a DBA and I'll die or pension as a DBA	- DBA role for life / no intention to change career paths
RD	Okay	
Q3.2	***… (hmm) … In terms of organizations being able to assist DBAs – what do you think … (hmm) … organizations can do to help DBAs get the needed skills for a big data environment?***	
R2	Well, I think it's more the DBAs responsibility to find the information than companies … (hmm) …or what vendors trying to inform people – so first of all companies that employs DBAs or going to use big data … (hmm) … should actually get their DBAs interested in big data technologies and then … (hmm) … sorry …	- Responsibility on DBA to do the research re BD - DBA must lead direction for BD technology stack by informing company - DBA to engage in self-directed learning
RD	So do you think it's all dependent on the direction of the company or their technology stack?	
R2	Yes, of course on the applications, if you have a company that embarks on a new business venture with massive data requirements then obviously they will have to consider that and re-skill their people	- Re-skilling towards BD is dependent on company direction
RD	Okay. Final question, … (hmm) … with the adoption of open source, how do you think the adoption of open-source technologies will assist DBAs in moving towards a role within big data? Do you think it will …	
R2	The same role it had when we have programs going open-source … (hmm) … I've actually not seen that – it would be very nice to have open-source technology available on the structure of the databases and how they work, you know … or in this case big data … machines or engines how they actually work in the background … maak dit sin?	- Likes the idea of open-source in order to explore internal dynamics of db software - Holistic/deeper understanding of BD technology stack
RD	Thank you – that concludes the interview … (hmm) … is there any questions you have for me or concerns?	

Romney Davids R2:3.1:DBA for life / no intention to change
Romney Davids R2:3.2:DBA responsibility to research big data
Romney Davids R2:3.2:DBA to lead technology direction
Romney Davids R2:3.2:Company must encourage big data research
Romney Davids R2:3.2:Re-skilling dependent on company direction
Romney Davids R2:3.2:Postive attitude towards open-source
Romney Davids R2:3.2:Explore database architecture through open-source

R2	No, nothing	
RD	Thank you very much	

APPENDIX F: Initial colour coding of codes

(This is an expanded list of some of the codes for interview questions 1.1, 2.1 and 3.1 to illustrate the process undertaken. The complete workbook is available upon request)

(Respondents R1 – R15 for interview question 1.1)

	A	B	C	D	E	F	G	H
1	**Respondent pseudonym**	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
2	**No of codes for respondent**	**11**	**6**	**8**	**30**	**6**	**9**	**17**
3	Transitioned from other roles	3	3	2	2	2	4	2
4	Tertiary qualification	1	1	1	0	1	1	1
5	Vendor certified	0	0	1	1	0	0	1
6	Diverse range of technical skills	1	0	1	3	0	1	6
7	Diverse range of non-technical skills	1	0	0	9	0	0	0
8	DBA role not chosen career path	3	0	1	3	2	0	3
9	Working experience	1	1	0	0	1	1	2
10	Wide range of responsibilities	0	0	0	0	0	0	0
11	Working environment	0	1	0	3	0	0	0
12	Miscellaneous	1	0	2	9	0	2	2
13		Leadership skills	Mainframe operator background	Tertiary qualification (National Diploma)	Development specialist background	Development specialist background	Development specialist background	Network Administrator background
14		Senior DBA experience	Development specialist background	Development specialist background	Consultant roles	System administrator background	OS specialist background	Unix OS skills
15		Development specialist background	Development to DBA transition	Transition from developer to DBA role	No DBA experience	Senior DBA experience	Environment specialist background	Development specialist background
16		Transition from Developer to DBA Role	Senior DBA experience	Learned on the job	Mentoring skills	Tertiary qualification (Degree)	Specialist DBA role	On-the-job learning
17		Oracle skills	Tertiary qualification (National Diploma)	Vendor certified DBA	Programming skills	Position aspired towards	Senior DBA experience	Unix OS skills
18		Development specialist background	Challenging environment	Trust relationship between development and DBA	Restricted working environment	Limited DBA opportunities	Tertiary qualification	SQL Server skills
19		Professional respect		Transitioned to fill knowledge gap	Creative mindset		Broad skillset required	Diverse database technology skillset
20		Remuneration		Programming skills	Trouble-shooting skills		Transition from Developer to DBA role	Self-taught MySQL skills
		Positon aspired towards			Experience is key		DBA career path	Senior DBA experience

Cat Descriptions | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | 3.1 | 3.2

	I	J	K	L	M	N	O	P
1	**R8**	**R9**	**R10**	**R11**	**R12**	**R13**	**R14**	**R15**
2	**7**	**5**	**4**	**7**	**12**	**15**	**14**	**18**
3	2	2	1	2	0	1	5	0
4	1	1	1	1	2	1	2	2
5	0	0	0	0	0	0	0	0
6	1	0	0	0	1	4	0	1
7	0	0	0	0	0	2	0	0
8	1	1	1	2	1	4	3	7
9	1	0	0	1	1	1	1	0
10	0	0	0	0	6	1	0	6
11	0	1	0	0	0	0	0	0
12	1	0	1	1	1	1	3	2
16	Transition from Developer to DBA role	Transition from System Administrator to DBA	University subject sparked DBA interest	Techie background	Security	Promotion to DBA	Application Support background	Intership
17	No intention to become DBA	Enjoys DBA role		Call centre agent background	Integrity of data	SQL Server skills	~~Network Engineering background~~	First job as Junior DBA
18	Senior DBA experience			DBA not by choice	Disaster recovery	Oracle skills	Network Administrator background	Moved jobs because of remuneration
19	Automated monitoring system			Tertiary qualification (National Diploma)	Compliance governance	Leadership skills	System Engineering background	Oracle skills
20					Database accessibility to end-users	Specify training requirements	6 years DBA experience	Assigned a mentor
21					Database stability	Design skills	DBA role through opportunity that arose	Learned DBA skills on the job
22					Oracle skills	Database architecture skills	DBA is a specialized role	Database availibility
23					Employment more important than being DBA	DBA not by choice	Management prospects	Database accessibility to end-users
					Dislike for programming	Bursary student	Customer Relationship	Monitor storage

Cat Descriptions | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | 3.1 | 3.2

(Respondents R1 – R15 for interview question 2.1)

Respondent pseudonym	R1	R2	R3	R4	R5	R6	R7
No of codes for respondent	11	13	16	11	5	20	9
Shortage of big data exposure	2	3	1	2	2	0	1
Limited big data knowledge	0	0	0	0	0	0	0
DBA role and responsibilities in big	1	6	4	2	1	3	3
Big data environment	6	2	5	0	0	7	0
Future career plans	1	2	4	4	0	0	4
Miscellaneous	1	0	2	3	2	10	1
	Unsure about future involvement in big data	Big data not new	No big data exposure	Questions whether size defines big data	Experience in large volumes of data	Companies don't understand big data	No big data exposure
	Administor large volume relational databases	Big data linked with volume	Big data technology the way forward	Exposure to 100TB RDBMS database	No business intelligence experience	Companies collecting data	Read about big data
	Volume not sole criteria for big data	No big data exposure	Companies analysis human behaviour	Data is everywhere	Experience in data warehousing	Using BI to understand data	~~Good all-round understanding of the concept~~
	No big data opportunities in company	Database structure should be de-normalized	Data storage is critical	Interested in problem-solving	No insight into big data environment	Companies aware of big data's potential	General understanding of big data
	Change jobs to get big data exposure	Complex environment	More dependencies with cloud	No interest in big data	Considers volume as big data	No clear big data direction	Not changing to big data
	Very similar role to RDBMS	No big data opportunities in company	Future outside of IT	No big data exposure		DBA potential advance to data architecture	Continue specializing within RDBMS
	Big data DBA has production focus	No big data experience	No interest in big data	Exposure in high volume databases		DBA potential advance to data management	RDBMS will continue to exist
	Continuous flow of data	Exposure to index-sequential files	No financial support from company for courses	Data useless if cannot be processed		Specialized environment/role	Not changing to big data
	Monitoring of data feeds	Index-sequential files is big data foundation	Courses very expensive	Interest in problem-solving		New roles emgerging	Big data is "real-world" problem
	24x7 data streams	Big data won't last	RDBMS & big data DBA role the same	No interest in big data		New challenges emerging	
	Production oriented environment	No insight into big data role	Specialist DBA roles	No big data experience		Specialist DBA roles	
		Boring environment	Specialist backups			Complex	

Cat Descriptions | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | 3.1 | 3.2

R8	R9	R10	R11	R12	R13	R14	R15
4	8	6	4	4	8	4	3
1	0	0	0	1	1	0	1
0	0	3	1	0	0	1	0
2	4	0	1	3	4	3	1
1	1	1	0	0	2	0	0
0	3	1	1	0	0	0	1
0	0	1	1	0	1	0	0
No big data exposure	Limited big data knowledge	Limited big data exposure	Companies moving towards big data	No big data exposure	Read about big data	Big data about data management	No big data exposure
Read up for interview	Refers to as on-relational data	Researched noSQL for client	Exposure dependent on client requirements	Only heard about big data	No big data exposure	Exposure dependent on environment	Upskill if client requires it
No insight into big data role	Limited big data technology knowledge	Exposure dependent on client requirements	Foresees future in big data	Don't want to comment on big data	Big data associated with 4th industrial revolution	Limited big data knowledge	Limited input re big data DBAs day
Similar routine checks as RDBMS	Interested in big data	Interested in big data	No insight into big data role	No comment on big data DBAs day	General understanding of big data	No insight into big data role	
	Time constraints limits research opportunity	Upskill via courses			Knowledge of big data tools within SQL Server		
	Transition if opportunity arises	Very similar role to RDBMS			Application DBA more exposed to big data		
	No insight into big data role				Production DBA to administer environment		
	Refers to possible "analytical stuff" – analytical type role				Data miners have advantage		

Cat Descriptions | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | 3.1 | 3.2

(Respondents R1 – R15 for interview question 3.1)

	A	B	C	D	E	F	G	H
1	**Respondent pseudonym**	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
2	**No of codes for respondent**	**14**	**9**	**11**	**18**	**5**	**11**	**13**
3	Skills development opportunities	2	1	3	0	3	1	3
4	Interest in RDBMS	0	4	3	0	0	0	1
5	Future career aspirations	1	1	4	0	1	2	1
6	Reasons for limited opportunities	2	0	0	2	0	1	4
7	Methods of upskilling	3	0	0	7	0	0	0
8	Ability to transition	3	0	0	0	1	1	0
9	Miscellaneous	3	3	1	9	0	6	4
10		No big data career plans in company	No career pathing towards big data	No career pathing towards big data	Companies must keep DBAs happy	No big data career plans in company	Career path determined by individual	No big data career plans in company
11		Big data DBA separate path to RDBMS	Enjoyment for job	Flat company structure	R&D important for DBAs	Perception nobody else working on big data	Training focused on specific solutions	Company scared for change
12		Decide which path to pursue	DBAs don't change career paths once settled	Long service in existing job roles	Allow for self-directed learning/experimentation	No company plans for big data	No holistic big data training available	No talk of big data in company
13		Can do both RDBMS & big data DBA work	Analyst background	No intention from employees to move	Skillset grows with practising	Has ability to transition to big data	Companies realizing big data potential	Only change technology if absolute must
14		Has ability to transition to big data	Business Analyst potential career path for DBA	Pay for own courses	Company size determines technology stack	RDBMS DBA in the future	Slow move towards big data	Companies don't understand benefits of big data
15		Re-skilling & re-certification required	DBA by choice	DBAs not interested in changing roles	Big data still busy evolving		Big data solutions bringing DBAs into contact with other departments	Cost implications potential prohibiting factor (staffing/re-skilling)
16		Industry popularity/size makes job prospects low	Enjoyment for job	No talk of big data in company	New tools not always more efficient		Companies realizing big data's potential/impact	Big data prospects dependent on company
17		RDBMS DBA knowledge/experience key to transitioning	Remuneration is important	Incorporate architecture into skillset	DBAs provide technical guidance		Architectural role in future	Infrastructure through to process change is needed
		Expand current skillset	DBA for life / no intention to change	Prefers to enhance development skills	DBA technical ability not always		Prefers problem-solving roles	DBA still in immediate future

Cat Descriptions | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | **3.1** | 3.2

	I	J	K	L	M	N	O	P
1	**R8**	**R9**	**R10**	**R11**	**R12**	**R13**	**R14**	**R15**
2	**2**	**7**	**6**	**3**	**4**	**6**	**6**	**4**
3	0	1	1	2	1	1	2	0
4	0	3	0	0	0	1	0	0
5	1	1	1	1	2	3	2	1
6	0	2	2	1	1	0	0	1
7	0	0	2	0	0	1	0	2
8	0	0	0	0	0	0	0	0
9	1	0	0	-1	0	0	2	0
10	Additional responsibilities for RDBMS DBA	Unaware of company big data plans	No big data requirement in company	No company plans for big data	Unaware of company big data plans	Unaware of company big data plans	Wanted to read up on big data tools	Company encourages certification
11	Possibly support big data environment	Big data dependent on company policy	Upskilling for big data through courses	Only able to upskill in private capacity	Upskill if company requirement	Upskill as vendor releases products	Limtied big data knowledge	Consulants must be certified
12		Company technology stack & plans determine DBAs big data prospects	Upskill on practice environments	Future in big data & security space	Future outside of DBA role	RDBMS will continue to exist	No comment on company big data plans	Upskill if company requirement
13		Future in big data	Single big data specialist within company		Future outside of IT	No thought on future prospects	Future as SQL Architect	Future outside of DBA role
14		Combination of RDBMS & big data DBA in future	Upskill on big data for customer support purposes			Future outside of DBA role	Help to design database solutions	
15		RDBMS & big data will be complimentary	Future career in security/ethical hacking			Studying alternate career path (digitilization)	Unaware of company big data plans	
16		RDBMS will continue to exist						
17								

Cat Descriptions | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | **3.1** | 3.2

APPENDIX G: Codes grouped by category

(The full list of codes for two categories for each theme is shown below. The complete workbook is available upon request)

(1) Theme - Skills and responsibilities (RDBMS environment)

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1.1, 1.3, 1.4	Operating System Knowledge	1		1			1	3	1	1	1	1	1		1	1	**11**	Diverse range of technical skills
1.1, 1.3	Database Vendor Skills	3	1			3		4	1		3		1	3	1	2	**10**	
1.1, 1.3, 1.4	Programming Knowledge		1		1	2	1	2	1		1					2	**8**	
1.1, 1.2, 1.3, 1.4	Diverse Skillset	1	1		1		2	3	1								**6**	
1.3, 1.4	Networking Knowledge							1		1	1	1	2			1	**6**	
1.1, 1.2, 1.3, 1.4	Trouble-shooting			1	1		4	1					1			2	**6**	
1.3, 1.4	Analytical Skills		1						1	1	1				1		**5**	
1.3, 1.4	Business Knowledge		2	1	2								1				**4**	
1.2, 1.3	SQL Skills	1	1			1									1		**4**	
1.3	General Technical Knowledge	1			1		1										**3**	
1.1, 1.3	Hardware Knowledge				1			1		1							**3**	
1.2, 1.3	Scripting Skills					1		1		3							**3**	
1.3	Storage Knowledge										1	1	1				**3**	
1.2, 1.3, 1.4	Understand Entire Solution				2			2					4				**3**	
1.3, 1.4	*Cloud Technology Knowledge*											*1*				*1*	***2***	
1.4	*Application knowledge*			*1*													***1***	
1.2, 1.3, 1.4	*Business Intelligence Skills*														*3*		***1***	
1.4	*Cross-skilling*				*1*												***1***	
1.4	*Database Management services to*														*1*		***1***	
1.1	*Design Skills*													*1*			***1***	
1.3	*Governance Knowledge*												*2*				***1***	
1.3, 1.5	*Mathematics*			*1*			*1*										***2***	
1.2	*Product Licencing*													*1*			***1***	
1.2	*Project Management Skills*										*1*						***1***	
1.4	*Solution Architecture knowledge*												*1*				***1***	
1.1, 1.2, 1.3, 1.4	Performance Tuning & Optimization		2	1		3	2	1		2	1	1		1		1	**10**	Wide range of responsibilities
1.1, 1.2, 1.3	Availibility	3		2		1	1	3					4	2		3	**8**	
1.1, 1.2	Backup and Recovery			1	1			2			1	1	2	1		3	**8**	
1.2	General Database Administration			1			1	1		1	2			1	1		**7**	
1.1, 1.2	Database Monitoring							1		1	1		2	3		1	**6**	
1.2, 1.3	End-user Support	1		1				1			1	1	1				**6**	
1.1, 1.2, 1.3, 1.4	Security						1		2			3	5	2	2		**6**	
1.2, 1.3	Customer Support			2							3				6		**3**	
1.2, 1.4	Developer Code Reviews				1							1			1		**3**	
1.3, 1.4	*Requirements Analysis*							*2*					*2*				***2***	
1.2	*Capacity Planning*						*1*										***1***	
1.1	*Compliance Governance*												*1*				***1***	
1.2	*DBA & developers responsibilities overlap*															*1*	***1***	
1.4	*Requirements Analysis*							*2*									***1***	

(2) Theme - Background and education

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1.1	Development background	3	2	2	1	1	2	1	2								**8**	Transitioned from other roles
1.1	Application / Desktop / Technical support										1	2		1	1		**4**	
1.1	System administrator background		1			1	1			2							**4**	
1.1	*Network Administrator background*							*1*							*1*		***2***	
1.1	*Consultant roles*				*1*												***1***	
1.1	*Customer Relationship Management background*														*1*		***1***	
1.1	*Environment specialist background*						*1*										***1***	
1.1	*Experience in various roles within IT*														*1*		***1***	
1.1	*System Engineering background*														*1*		***1***	
1.1	Tertiary qualification		1	1		1	1	1	1	1	1	1	1	1	1	1	**13**	Tertiary qualifications
1.1	*Multiple tertiary qualifications*												*1*		*1*		***2***	
1.1	*Incomplete tertiary qualification*	*1*															***1***	
1.1	*Self-taught*				*1*												***1***	

(3) Theme - Training and upskilling

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1.3, 1.4, 1.5	On the job experience	1		2	1	1		3	1					1		1	**8**	Importance of on-the-job experience
1.3, 1.5, 2.4, 3.1	Self-directed learning	1		2	2			2	2	1	2	1			1	1	**10**	Methods of upskilling
1.3, 2.3, 3.1	Training courses			1					3	2	2	2				1	**6**	
1.3, 3.1	Certifications	2			2					1			1			1	**5**	
1.3, 2.4	Awareness of emerging technologies	1	1				2	1									**4**	
1.1, 1.3, 1.5, 2.3, 2.4	Internet-based content	1			3	3	2										**4**	
1.3, 1.5, 2.1, 2.3, 2.4	Lack of company sponsored courses			4		1						1					**3**	
1.3, 3.1	New vendor product releases												1	1		1	**3**	
1.3, 2.4, 3.1	Practice environments				3						3		1				**3**	
1.3	*Upskill in private capacity*	*1*			*1*												***2***	
2.3	*Attend seminars*				*2*												***1***	
1.3	*Continuous upskilling required*															*1*	***1***	
2.3	*Database management challenging*					*1*											***1***	
1.3	*DBA skill is developed not taught*													*1*			***1***	
1.5	*DBA skills can be learned/taught*															*1*	***1***	
3.1	*DBAs to grow development skills*				*1*												***1***	
2.3	*Formal qualification improves prospects to*			*1*													***1***	
1.5, 2.4	*Knowledge sharing with DBAs*					*2*											***1***	
1.5	*Lack of knowledge sharing*		*1*														***1***	
1.3	*Learn as required*				*1*												***1***	
2.3	*Learn new tools*	*1*															***1***	
2.3	*Limited opportunity given to learn*									*1*							***1***	
2.3	*No single big data solution*						***1***										***1***	
2.4	*Requirement to upskill in some big data technologies*					*1*											***1***	
2.3	*Re-skilling possibly complexed*						***1***										***1***	
2.3	*Solution comprises multiple vendors*						***1***										***1***	
1.3	*Training courses not primary means of*	*1*															***1***	
2.4	*Upskilling method same for RDBMS and big data*						*1*										***1***	

(4) Theme - Adoption of big data

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
2.1, 2.4	No big data exposure	1	3	1	2			2	1				1	1		1	**9**	Shortage of big data exposure
2.4	Limited exposure with big data tools	1	1							1	2						**4**	
2.1, 2.4	No big data opportunities in company	1	2					1									**3**	
2.1	*Exposure dependent on client requirements*										*1*	*1*					***2***	
2.1	*Limited big data exposure*					*1*					*2*						***2***	
2.1	*Change jobs to get big data exposure*	*1*															***1***	
2.1	*Experience in large volumes of data*					*1*											***1***	
2.1	*Exposure dependent on environment*														*1*		***1***	
2.4	*Older DBAs less involvement in big data*			*1*													***1***	
2.1	*Received big data training*														*1*		***1***	
2.1, 2.4	Limited big data knowledge							2	2	2		1	1	1	1		**7**	Limited big data knowledge

(5) Theme - Skills and responsibilities (Big data environment)

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
2.2, 2.3	Data Analytics		1	1		1	1				2						5	Expansion of technical skills
2.2, 2.4	Cloud technologies						1					1		1	1		4	
2.3	Data Mining		1	1											1		3	
2.3, 2.4	Focus on data					1		1						3			3	
2.3	Programming skills							1			1			1			3	
2.2, 2.3	*Business knowledge*		*1*				*1*										***2***	
2.2, 2.3	*Learn new big data toolsets*	*1*					*1*										***2***	
2.2, 2.3	*Process optimization*					*2*	*1*										***2***	
2.2, 2.3	*Understand entire solution*							*1*						*1*			***2***	
2.3	*Additional non-traditional DBA skills will be*							*1*									***1***	
2.2, 2.3	*Big data database management skills*										*2*						***1***	
2.3	*Big data query languages*										*1*						***1***	
2.3	*Business Intelligence skills*									*1*							***1***	
2.2, 2.3	*Cross-skilling*							*2*									***1***	
2.2	*Different skillset required*							*1*									***1***	
2.2	*Infrastructure knowledge*						*2*										***1***	
2.3	*Large datasets*				*1*												***1***	
2.3	*Non-relational databases*											*1*					***1***	
2.3	*Operating system skills*							*1*									***1***	
2.2	*Project management skills*							*1*									***1***	
2.2, 2.3	*Scripting skills*													*2*			***1***	
2.2	*Security*										*1*						***1***	
2.2, 2.3, 2.4	Less intervention from DBA		4	1	2	1					1				2	2	7	Less technical skills
2.2	*Existing libraries make things easier*				*1*												***1***	
2.2	*Focus now on understanding &*				*1*												***1***	
2.2	*Less development from scratch*				*1*												***1***	
2.4	*More restricted environment*			*1*													***1***	

(6) Theme - Big data career pathing

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Categories
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
3.2	Company sponsored courses	1		1		2						1					4	Skills development opportunities
3.2	Concern around training courses									4		1	2				3	
3.2	*Company must promote big data research*		*1*										***1***				***2***	
3.2	*Company support with non-dba related skills*							*1*									***1***	
3.2	*Online courses*									*2*							***1***	
3.2	*Practical exposure important*												*1*				***1***	
3.2	*Provide environments to practice*												*4*				***1***	
3.2	*Unsure who's responsible for upskilling*														*1*		***1***	
3.2	*Upskill on big data tools*														*2*		***1***	
3.2	*Upskill through collaboration between teams*							*1*									***1***	
3.2	*Upskilling programmes already in place*													*2*			***1***	
3.2	*Upskills after hours*									*1*							***1***	
2.1, 3.1	Future outside of DBA role (other IT careers)			4			1	2			1		2	2	1	1	8	Future career aspirations
3.1	Architecture role			1			1								1		3	
2.4, 3.1	RDBMS DBA	1	6			1											3	
3.1	*Future within big data space*									*1*		*1*					***2***	
3.1	*No thought on future prospects*													*1*			***1***	
3.1	*Possibly support big data environment*								*1*								***1***	

(7) Theme - Transitioning to a big data role

IQ	Codes belonging to category	Respondents from which codes emerged (R1-R15)															Strength of code (N)	Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
2.3	Re-skilling required		1	1	3	1	1							2		1	7	Re-skilling
2.3	No re-skilling required							1				1			2		3	
2.3	*Expand skillset*	*1*															***1***	
2.3	*Transition to Data Engineer*	*1*															***1***	
2.3, 3.1	DBA has skills to transition	4			1	1	1			1	2	2				2	8	Ability to transition
2.3	*Transition dependent on tools used*	*1*													*1*		***2***	
2.3, 3.1	*Transition influenced by experience*	*1*	*2*														***2***	
3.1	*Can do both RDBMS & big data DBA work*	*1*															***1***	
2.3	*Database exposure will assist transition*					*1*											***1***	
2.3	*DBA has some skills to transition*						*1*										***1***	
2.3	*Learning curve when transitioning*					*1*											***1***	
2.3	*Similar field as RDBMS*					*1*											***1***	

APPENDIX H: Respondent quotes used in thesis

(The coded transcript from which these quotes were extracted for all respondents is available upon request)

Quote No.	Quote extracted from interview transcripts	Respondent who made the quote	Page no. in transcript
1	*"…you need some kind of OS skills, for us most of our work is done in Linux or Unix so those, the OS skills are important…"*	R1	5
2	*"…When it comes to the networking side, I think you know sort of on a small sort of you know I know what IP addressing is…"*	R7	26
3	*"…(Hmm) … Ja I think knowledge of the type of the database that they using … (hmm) … they have to have pretty good knowledge of that…"*	R5	63
4	*"… making sure that it is running efficiently, if there is any bottle-necks that you sort of get to it that you can find out what it is and what is causing it … (hmm) … most cases it's users running exceptionally bad SQL that's causing the problem …"*	R6	76
5	*"…It becomes a bonus if you have lets say programming skills because sometimes the developers or the client will say that the database is slow and maybe they getting … an error …"*	R15	205
6	*"…communication is very important because a lot of the times people going to ask you okay what went wrong then you need to describe to them not in a technical DBA type of view-point but for their understanding…"*	R3	31
7	*"…people skills … you need to be you need to be able to communicate your soft skills, you need to be able to communicate effectively with with other people…"*	R9	125
8	*"…I think you got to have very good problem solving skills cause lot of the day is solving problems. (Hmm) … you got to be self-motivated…"*	R1	5
9	*"… (hmm) but I would say as a corporate citizen, I mean you work in a corporate environment … time-management is a must for any person on that basis …"*	R6	79
10	*"…and then the next thing they after I think is the performance tuning aspect … (hmm) … because we running such a lot of batch work in the background, especially at night, we need optimal performance to, because the batch window is about 7 hours long, by that time you should have everything done already – if there's any delay it'll have a major impact on the next day for the normal end users…"*	R3	27
11	*"I'd say to keep the database running for one, and running optimally so queries don't take too long …*	R6	61

	(hmm) … so that people can get their reports out … (hmm) … and making sure that everything is just running efficiently."		
12	*"… ja we do backups … our performance tuning … and then .. (hmm) … we do … (hmm) … like auditing as well on our databases …"*	R11	148
13	*"…if there's anything suspicious then I will raise a request for that and this is like for auditors purposes…"*	R12	160
14	*"…and then after hours that's then maybe if we identify there's a … a release or a change that needs to happen that will impact the system or the ability of the system then we push those for after hours …"*	R13	174
15	*"…we check for any file systems , tablespaces that have filled up … (hmm) … the previous night's backup that went through so that we know we can recover until the exact point-in-time..."*	R10	134-135
16	*"…and we obviously you need that skill to vet the code and see if it's relevant and if something fail you can actually fix it that time especially before production however if it fail you follow the protocol, engage the developer, discuss it with the guys and ja …"*	R14	188-189
17	*"I'm really lazy in the respect that email or if I can run a couple of scripts in the background that saves me 25 minutes of mundane boring work, I'd do that."*	R4	45
18	*"…No non-DBA related functions – only other function that could be considered perhaps non-DBA is checking software strategy & compatibility of software against the database version…"*	R8	116-117
19	*"… I think that is something that is necessary as a DBA … as a DBA I think it's necessary that you cannot have like you know sort of the purest type of DBA …. purest? … maybe it's not the …"*	R7	98
20	*"… so within the team we have like your teams that works with operating systems, storage, networks so when we work on the solutions those teams are integrated they work in the form of a project to deliver the solutions so the OS will build your operating systems, the storage will assign the storage and the network team will make sure your the servers you are able to access them…"*	R13	175
21	*"… ja it is … because we deal with the … (hmm) … the application team … the project management team … (hmm) … the change management team on a daily basis … so …"*	R11	149
22	*"…so there's a responsibility that you carry with you, and I find as a DBA in my perspective that's your biggest strength…"*	R4	42
23	*"…I know that some of the services be it financial be it industrial be it whatever they have to report to a certain body and if they cannot provide the reports at a certain time they gonna be fined, there gonna be penalties so as a DBA I need to know that information so that I have to make sure that*	R12	162

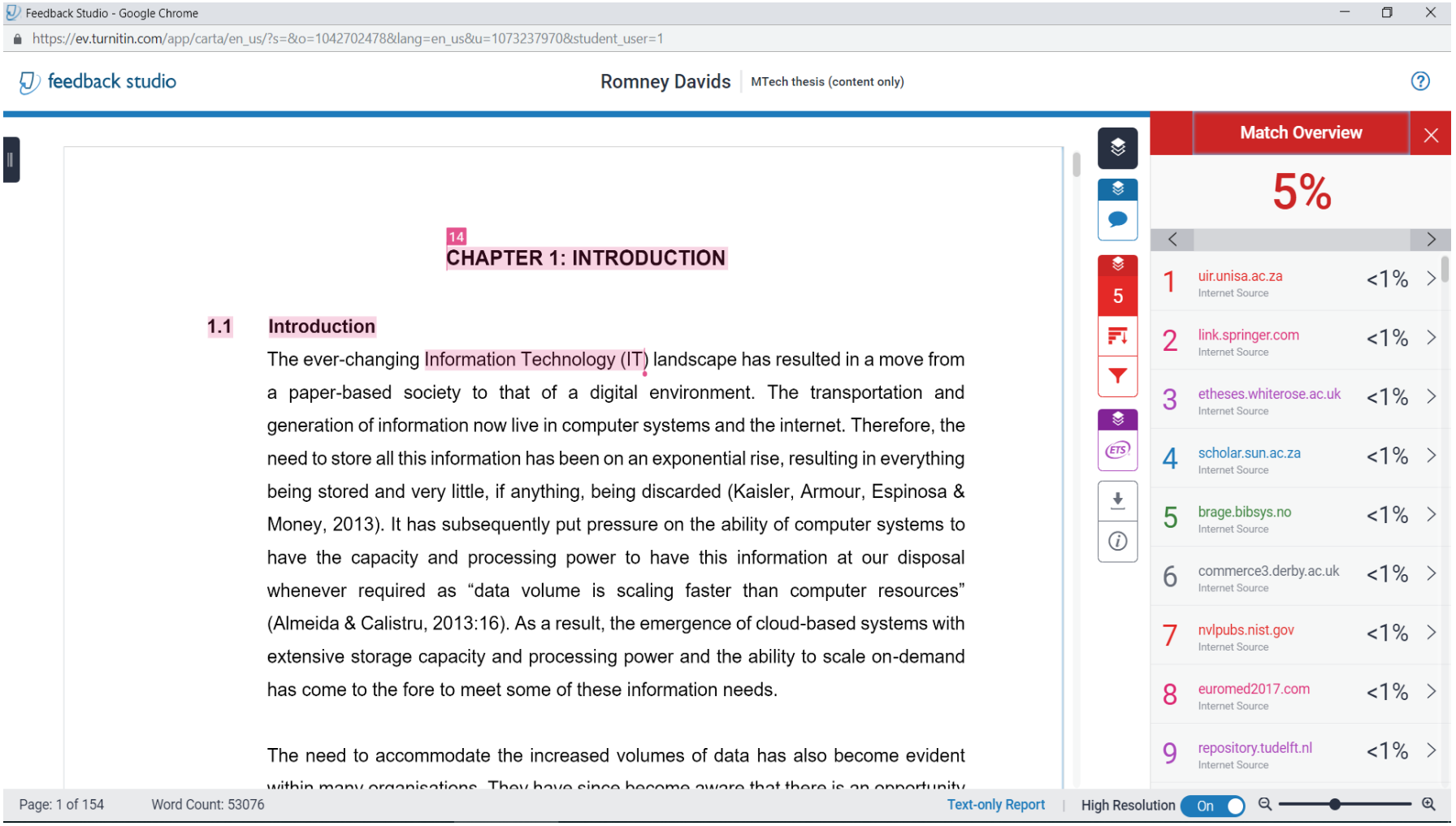
	their system it's up and running…"		
24	*"…Okay, … (hmm) …, I started off in programming really, … (hmm) … then moved to Systems Admin and more recently probably the last 6 years I've been doing DBA work…"*	R5	60
25	*"… my background started when I completed … (hmm) … my studies … (hmm) … round about 2007 and then I joined the company as a junior IT specialist whereby in that time we just rotated across different sections you know just to learn about the operations in the organisation…"*	R13	172
26	*"… ok ja so I've fulfilled a couple of roles in IT sector … I think I started in IT 2000 … 2006 and 7 … 2006 going into 2007 ja … so I did mostly application support, network engineering, network administration during that time … (hmm) … later most to [CompanyNameWithheld] … [NameWithheld] … support satellite, virtual satellite support which also involves a lot of networking as well and later moved to … later moved to another company where I did some application support as well, moved to system engineering and then in 2012 moved to … (hmm) … moved to database management systems …"*	R14	185-186
27	*"…Okay … so I started … (hmm) … 25 years ago I studied BCom at [UniversityNameWithheld] … I did a BCom Information Systems and then from there I worked my way up to System Admin..."*	R9	122
28	*"…I think if I studied a bit more, stayed a bit more current I could probably commanded better, better salary … (hmm) … and possibly might have opened different opportunities like I'm seeing now in the BI and DSS space…"*	R4	42-43
29	*"…No … that was not my intention … (hmm) … actually IT was not my intention … (giggles) … so basically I was studying to get money to be employable and survive …"*	R15	202
30	*"…(Hmm) …, so I started off as a developer and when I was working as a developer, the DBA guys always seemed to be very knowledgeable and well paid, so I always thought it was a good position to aim for…"*	R1	2
31	*"…so I think the qualification prepares you to be able to be an independent thinker like you know like a person who is able to be given a problem and then you able to supply solutions yourself…"*	R13	177-178
32	*"…Okay, I work with a guy that has got 20 years of oracle database experience with no qualifications - except matric and he's doing exceptionally good - he's our application production DBA…"*	R2	16
33	*"… (hmm) … a formal qualification is very important within our field … (hmm) … I've done a national diploma in IT but I've decided not to do it further so that I could do my vendor training instead because I felt that it was a bit more important to … (hmm) … than going further on my studies because within Oracle there's a lot of*	R10	139

	courses say … (hmm) … with each technology you will have to go on a course …"		
34	*"… puts you in a better position than when you have a degree because then you actually get into contact with the with the technical stuff immediately …"*	R9	127
35	*"… so I always say to the guys that go and write those exams and come and say I have the certification, I'll say that you have a certification but do you have the knowledge of how to use the tool itself and then most of the time it's a contrast you know in a sense that a person go get a certification but when comes to applying the knowledge …"*	R13	178
36	*"…what has happened in real world you actually knocked your head, you actually fallen and you picked yourself up, you actually recovered from a disaster, you know those type of things … you (some words inaudible) migrated from one system to another, and those type of things so for me experience …"*	R7	104
37	*"You just need to read … that's all basically …"*	R11	150
38	*"… besides the certification … besides going onto training … you need to do stuff hands-on so that ja … with no pressure because most of the time either if it's project related there will be time-lines and there are costs involved which understand you end up being under pressure you know to deliver something that sometimes you don't even knowledge on you just go and google stuff and you understand what I'm saying so I believe a lab environment would …"*	R12	170
39	*"…I mean currently I'm doing also part-time SQL DBA work but I'm not certified yet but the company is not going to send me to do that but I don't think I will pay but maybe in the future but not now …."*	R3	32
40	*"…I don't see myself going with what is now becoming known as big data unless I move to another company where, that's big data…"*	R1	6
41	*"…Yes … I am involved in that to a degree at the moment (hmm) because we do a lot of archiving, we keep tons of data …"*	R5	65
42	*… I've looked at … (hmm) … like I said I've looked at Hadoop … (hmm) … when there was a requirement last year …*	R10	143
43	*"Read up to understand for the interview"*	R8	119
44	*"… (hmm) … it is something that I would like to actually focus on … something that I would actually like to specialise on. RDBMS world or I mean non-Big Data sort of environments are not going to die yet, is not going to be dead yet."*	R7	106
45	*"…I wouldn't, well I haven't and I wouldn't to rush off to become a big data, to drop everything and learn all the big data toolsets – but definitely there are some toolsets that are standing out from the pack like Hadoop and Apache Spark and R I think is one of the other big ones – so there are definitely now becoming – it's becoming clear what*	R1	8

	the tools are that are going to move forward – so if I was moving that way then definitely there are things you can follow up on – that answer your question…"		
46	*"…Well, ja, I've never worked in any big data environments so I think from my personal view-point that is definitely going to be the future…"*	R3	33
47	*"…it's fascinating to me I just wish I more time to go and and look into it …"*	R9	128
48	*"…Not yet, but it is something that I will be looking at … I mean I have some ideas of my own in terms of how to implement some of the solutions …"*	R6	87
49	*"…I would actually like to specialise on. RDBMS world or I mean non-Big Data sort of environments are not going to die yet, is not going to be dead yet…"*	R7	106
50	*"…Yes of course, ja data mining … (hmm) … again analytical skills I think will be more important than what it is currently for production type DBAs…"*	R2	20
51	*"…Coding definitely is … (hmm) … I think as much as we not coding the applications but we are coding the data.."*	R13	181
52	*"…I think on the BI stuff, a Big Data DBA will you know sort of … I'm just going to use an example to provide reports on a daily basis of so much data … he needs to be able to actually understand how to get those reports … (hmm) … how to get the data … how to manipulate the data and how to present the data … so definitely yes…"*	R7	110
53	*"…Ja … at the moment we've been very very … (hmm) … working closely … basically it's cloud computing … it's the one that has been … (hmm) … the talk of our town because that's where the world is going basically … environment … you no longer going to be managing your own environment, we'll all be hosted in the cloud computing so that's the one basically that we've been looking at it closely ja … "*	R14	196
54	*"… you would have to upskill the way you query the databases … (hmm) … even though you know the concept of the varying database …"*	R10	143
55	*"…Most definitely yes … ja … because most of the things were manual and now things are automatic it will be just a matter of a click of a button …"*	R15	210
56	*"…I don't know, I've got no idea – maybe very boring … (hmm) … maybe that type of database administrator will also have to be more business type, user, or business person than a technical DBA…"*	R2	18
57	*"…I'm not sure … I don't want to lie …"*	R11	153
58	*"… (hmm) … I would imagine it would be similar to what we do … they would also have to have backups … (hmm) … even though big data environments are highly redundant … (hmm) … I would also imagine it would … (hmm) … you'd have to check backups and all the things we do …"*	R10	141

59	*"…Well I think they could … (hmm) … pay for courses and things like that, I mean that would definitely help … (hmm) … like more formalised courses … (hmm) … and Ja, that would help …"*	R5	70
60	*"… besides the certification … besides going onto training … you need to do stuff hands-on so that ja … with no pressure because most of the time either if it's project related there will be time-lines and there are costs involved which understand you end up being under pressure you know to deliver something that sometimes you don't even knowledge on you just go and google stuff and you understand what I'm saying so I believe a lab environment would …"*	R12	170
61	*"… (hmm) … I'm thinking … in South Africa I specifically now I think they can still … they they … they can profit to start using online training because a lot of the companies still send their people out you know to training centres or whatever … training environments … then they have to stay away from work for a week …"*	R9	132
62	*"…Okay, where we are you just a DBA and that's it – there's no career path …"*	R2	22
63	*"…(Hmm) …, well I can only speak for my organisation, I don't think they have really plans for that, I mean it's more kind of by default you know, nobody else is doing it so you know, I'll do it kind of thing…"*	R5	69
64	*"…(hmm) … I don't have that information … ja … maybe [PersonNameWithheld] the one before me might have that information ja …"*	R14	197
65	*"… it's a longer term type of thing, they need to understand that you know right at the end that all the systems are going to be switched off unless we actually go to Big Data … (hmm) … prior to that it's the … it's it's sort of why do we need to go to Big Data … ? These are new technologies then we going to have to get like new you know sort of staff … you have to … (hmm) … redo all our infrastructure, those type of things …"*	R7	113
66	*"…companies are realizing the importance thereof so they are moving towards it, maybe not as fast as they should…"*	R6	88
67	*"…If I'm still going to be in IT, I will, one of my things is rather to move more to the development environment, maybe adding a bit of other languages in, maybe a bit of architecture…"*	R3	38
68	*"no … mostly more SQL architect …"*	R14	197
69	*"…(Hmm) … Not data big data as you've got it – I was actually looking at data warehouses…"*	R2	21
70	*"…Most definitely yes … ja … because most of the things were manual and now things are automatic it will be just a matter of a click of a button …"*	R15	210
71	*"…there would be branching into may different aspects and there would definitely be re-skilling required…"*	R6	85
72	*"…Yes, I feel that … well not formally … I don't*	R7	108

	think that the the the traditional DBA … (hmm) … sort of … (hmm) … role is that much of a threat yet, whereby you do have go learn something else right now, however it will get there…"		
73	*"…Certification obviously…"*	R2	19
74	*"…I think you'll definitely have to re-certify … ja … but also some of the … (hmm) … some of the … (hmm) … business intelligence stuff … (hmm) … is part of the DBA track so you'll probably be able to use a lot of that already … (hmm) …"*	R9	129
75	*"… because you write a certification first without … without even worked on that thing … then after the certification you start working on the thing … experiences the thing so according to me it should be the other way around … you should first work on the thing and then just get a certificate …"*	R15	210



Screen shot from TurnItIn.com