

*Full Length Research Paper*

# The use of big data analytics in the retail industries in South Africa

Matthew Ridge, Kevin Allan Johnston\* and Brian O'Donovan

University of Cape Town, South Africa.

Received 29 May 2015; Accepted 9 September, 2015,

The majority of publications around big data analytics are centred on technical algorithms or systems development. While research has been conducted on the use of big data analytics there is an apparent lack of studies which focus on industry-specific usage in South Africa. The purpose of the study was to assess the usage of big data analytics in the retail industries in South Africa. The benefits of using big data analytics are not specific to a particular industry. Retailers, for example, can use big data analytics to gain new insights about their customers in order to inform decision making around pricing and marketing. The usage of big data analytics was assessed by collecting data from interviews with retailers, big data vendors and professional services. The main finding of the study was that South African retailers are not using big data analytics. Some retailers are, however, using big data analytic platforms to improve the speed of processing large amounts of structured data and to deliver information cost effectively. The findings show that South African retailers find it difficult to identify a use case to justify the investment required for using big data analytics.

**Key words:** Big data, data analytics, retail industry.

## INTRODUCTION

Big data is often described as a collection of large and complex datasets which are difficult to capture, store, manage and analyse effectively using current database management software and concepts (Fan and Bifet, 2013; Kaisler et al., 2013). Big data is not a new concept, and companies such as Ebay, LinkedIn, and Facebook have been collecting big data since the mid-2000s (Davenport, 2013). Today a number of companies are collecting and processing large amounts of data on a daily basis. Google processes about 24 petabytes of data every day (Davenport et al., 2012), and retailers such as Walmart collect more than 2.5 petabytes of data every hour from customer transactions alone (McAfee and

Brynjolfsson, 2012). With the overwhelming amount of data being generated at a terabyte and even a petabyte scale, there is a need for big data analytics to obtain insights from big data (Chen et al., 2012; Singh and Singh, 2012). Big data analytics refers to a collection of analytic techniques and technologies which have been specifically designed to analyse big data to inform decision making (Fisher, et al., 2012; Kwon et al., 2014; Russom, 2011). Businesses can use big data analytics to improve target marketing, obtain additional business insights and to detect fraud (Manyika et al., 2011; Russom, 2011). The benefits of big data analytics are not limited to a specific industry, and virtually any firm in any

---

\*Corresponding author. E-mail: kevin.johnston@uct.ac.za.

industry can exploit the possibilities associated with big data analytics (Davenport and Dyché, 2013).

In the retail industry, for example, retailers can use big data analytics to gain new insights about customer behaviour and to improve decision making (Russom, 2011; Singh and Singh, 2012). Despite the potential benefits that research has attributed to the use of big data analytics, organisations have been slow to integrate big data platforms into their decision making framework (Kwon et al., 2014; Russom, 2011).

Much of the research on big data analytics has been centred on technical algorithms or system development (Kwon et al., 2014). Research has been conducted on amongst others the usage of big data analytics to understand customer relationships and experience (Spiess et al., 2014; Yadav and Kumar, 2015). While research has been conducted on the use of big data analytics by organisations such as The Data Warehousing Institute (Russom, 2011), there is an apparent lack of studies which assess industry-specific usage of big data analytics in South Africa. In response the objective of this study was to assess the usage of big data analytics in the retail industry in South Africa. The South African retail industry is a large and diverse group which includes clothing, furniture and grocery stores (Brown and Russell, 2007). The South Africa retail industry was chosen as it is the “largest in the sub-Saharan region and is globally ranked as the 20th largest retail market in the world” (Strydom, 2015, p. 465).

The research questions which emerged from the literature review were:

RQ1 How are South African retail companies defining big data and big data analytics?

RQ2 To what extent are South African retail companies using big data analytics?

RQ3 What value can South African retail companies gain from using big data analytics?

RQ4 What are the barriers for South African retail companies in implementing big data analytics?

RQ5 What techniques and technologies are South African retail companies using for analytics and big data?

RQ6 What vendor products are available to South African retail companies?

The structure of the paper is as follows. The next section reviews the current literature on big data analytics. The following section discusses the research methodology used followed by the data analysis and then a discussion on the findings of the study, closing with the conclusion.

## LITERATURE REVIEW

Many IT vendors and solution providers use the term ‘big data’ as a synonym for ‘more insightful data analysis’ (Davenport et al., 2012). Some regard big data as

propaganda used to sell Hadoop based systems (Fan and Bifet, 2013). However, the term big data is more meaningfully applied to a collection of large and complex datasets which are difficult to capture, store, manage and analyse effectively using current database management software and concepts (Fan and Bifet, 2013; Kaisler et al., 2013; Manyika et al., 2011). Big data not only creates difficulties because of the volume of data but also because of the variety of sources and its volatility. Consequently big data is characterised by what is termed the three V’s – high volume, high velocity and high variety. These require more advanced technologies and innovative processing in order to provide information of value for decision making (Chen and Zhang, 2014; Kaisler et al., 2013).

The volume of data refers to the size of the datasets which are collected (Chen and Zhang, 2014; Kaisler et al., 2013). The size of the datasets can be quantified in many ways such as the number of records, transactions, tables and files (Russom, 2011). Kaisler et al. (2013) quantifies big data in terms of volumes of data in the range of  $10^{18}$  exabytes and beyond. Generally, however, the size of datasets are defined in terms of terabytes or petabytes and most definitions do not assign an exact number to the volume of data (Russom, 2011; Singh and Singh, 2012). Data volume refers to the size of the datasets, such as the number of records or transactions, which are being stored and analysed by retailers. Retail organizations are used to working with large volumes of data but what makes big data different is the combination of volume with variety and velocity.

Big data can be obtained in a variety of data types from a variety of data sources (Chen and Zhang, 2014; Fan and Bifet, 2013; Russom, 2011). The difficulty is in integrating large amounts of data obtained in the form of: structured data (data stored according to fields in spreadsheets and relational databases), unstructured data (raw data such as text, video, audio and images) and semi-structured data (data which contains tags and other markers for separating data elements such as XML and RSS feeds) (Manyika et al., 2011; Russom, 2011).

Organisations have been collecting data from a variety of different data sources for many years but have not been tapping into all of their possible data sources. What has changed is that organisations are now starting to analyse and tap into this varied data (Russom, 2011). Data variety refers to the various data types which retailers are storing and analysing from data sources such as RFID tags, clickstreams and customer transactions (Manyika et al., 2011). Velocity of data refers to the rate at which data is generated and the rate at which data is processed (Chen & Zhang, 2014; Russom, 2011). Streams of data are being generated from a variety of sources and big data technologies allow for data to be collected, stored, retrieved and processed in real time (Fan and Bifet, 2013; Russom, 2011). In the context of retailers, the data velocity refers to the rate at which data

is created and extracted from applicable data sources in real-time or in near real time (Singh and Singh, 2012).

### **Big data analytics**

The analysis of big data to gain insights is a new concept (Russom, 2011). Big data analytics has been defined in a number of ways and there appears to be a lack of consensus on the definition. Big data analytics has been defined in terms of the technologies and techniques which are used to analyse large scale complex data to help improve the performance of a firm (Kwon et al., 2014). Russom (2011) defines big data analytics as the application of advanced analytic techniques on big data sets. Fisher et al. (2012) define big data analytics as a workflow which distils terabytes of low value data down into more granular data of high value. For the purposes of this paper, big data analytics was defined as the application of analytic techniques and technologies to analyse big data in order to obtain information which is of value for making decisions (Fisher et al., 2012; Kwon et al., 2014; Russom, 2011).

Analytics is a broad term which can be used to cover decision making which is data driven (Fisher et al., 2012). Analytics is also described as a process of developing actionable insights from the application of statistical models to problems and the analysis of existing and or simulated data (Cooper, 2012). In terms of big data, analytics has been defined as complex procedures which run over large scale datasets in order to extract useful information (Cuzzocrea et al., 2011). Analytics was defined as the process of extracting actionable information from large scale data in order to provide insights to drive decision making (Cooper, 2012; Cuzzocrea et al., 2011; Fisher et al., 2012). The main categories of analytics are descriptive, predictive and prescriptive analytics (Camm et al., 2014; Delen and Demirkan, 2013; Kaisler et al., 2013).

### **Descriptive analytics**

Descriptive analytics is the set of techniques which are used to describe and report on the past (Camm et al., 2014; Davenport, 2013). Retailers can use descriptive analytics to describe and summarise sales by region and inventory levels (Camm et al., 2014). Examples of techniques include data visualisation, descriptive statistics and some data mining techniques (Camm et al., 2014).

### **Predictive analytics**

Predictive analytics consists of a set of techniques which use statistical models and empirical methods on past data in order to create empirical predictions about the future or determine the impact of one variable on another (Camm et al., 2014; Shmueli and Koppius, 2011). In the

retail industry, predictive analytics can extract patterns from data to make predictions about future sales, repeat visits by customers and likelihood of making an online purchase (Camm et al., 2014; Shmueli and Koppius, 2011). Examples of predictive analytic techniques which can be applied to big data include data mining techniques and linear regression (Camm et al., 2014).

### **Prescriptive analytics**

Prescriptive analytics uses data and mathematical algorithms in order to determine the best course of actions to take based on a set of requirements and with the objective of improving business performance (Camm et al., 2014; Delen and Demirkan, 2013). Retailers can use prescriptive analytics to determine price mark down models to aid in setting discount levels to maximise revenue (Camm et al., 2014). Examples of prescriptive analytic techniques which can be applied to big data include optimisation methods (Camm et al., 2014; Delen and Demirkan, 2013).

### **Value of big data analytics in retail**

Initially, technological and economic factors limited the leveraging of big data (Devlin et al., 2012). Data which was initially omitted can now be included as big data platforms provide support for a high variety, velocity and volume of data (Devlin et al., 2012). The value of big data analytics is in the new insights which can be obtained from analysing big datasets in order to drive decision making (Fan and Bifet, 2013; Kaisler et al., 2013; Russom, 2011). Retailers can use these insights to optimise processes along the value chain (Manyika et al., 2011; Mohamed et al., 2012). Some of the functions retailers can apply big data to include price optimisation, customer micro-segmentation and marketing, inventory management, customer sentiment analysis and in-store behaviour analysis (Manyika et al., 2011; Mohamed et al., 2012).

### **Price optimisation**

Retailers can use a variety of data sources to help inform pricing decisions (Manyika et al., 2011). Retailers can take advantage of the granularity of data available on sales and pricing by performing an analysis to determine how the market demand is affected by certain price and product changes (Mohamed et al., 2012; Manyika et al., 2011). Insights into pricing can then be derived in order to facilitate an optimal pricing decision (Mohamed et al., 2012).

### **Customer micro-segmentation and targeting**

Retailers have access to high volumes of data for

segmentation from a wide variety of data sources such as loyalty programs, location data and clickstream data (Manyika et al., 2011; Mohanty et al., 2013). The increasing sophistication of analytics allows retailers to integrate and analyse customer data in order to divide customers into more granular micro-segments (Manyika et al., 2011). Customers can be segmented by individual behaviours by using big data analytics to analyse data collected on customer behaviour at each touch point (Madsen, 2013; Manyika et al., 2011). By viewing customers at an individual level, retailers can personalise and tailor marketing techniques (e.g.: product recommendations) to increase customer satisfaction levels (Mohamed et al., 2012; Ularu et al., 2012).

### ***Inventory management***

The use of big data analytic tools can help retailers improve their inventory management (Manyika et al., 2011). Retailers can improve their stock forecasting by combining multiple datasets, such as sales histories and seasonal sales, and using analytics to predict changes in demands (Mohamed et al., 2012; Manyika et al., 2011). Retailers can analyse stock utilisation data from data sources such as bar code systems to help automate stock replenishment decisions thereby reducing incidents of stock delays (Mohamed et al., 2012; Manyika et al., 2011).

### ***Customer sentiment analysis***

Sentiment analysis leverages the large volumes of data generated from customers on various forms of social media to help inform decision making (Manyika et al., 2011). Retailers can use sentiment analysis to monitor real time responses to marketing campaigns and can adjust processes accordingly (Davenport et al., 2012; Manyika et al., 2011).

### ***In-store behaviour analysis***

Retailers can leverage a number of technologies, such as real time location data from smartphones, to collect information on customers' in-store behaviour (customer footpath and time spent in different parts of the store) (Manyika et al., 2011). The information collected on in-store behaviour can be analysed to derive insights on improving certain aspects of a retailers store such as store layout, shelf positioning and product mix (Manyika et al., 2011; Mohanty et al., 2013).

### **Barriers to using big data analytics in retail**

Due to the complexity of big data there are a number of barriers related to analysing big data (Rajakumar and

Sushma, 2013). Some of the main barriers which arise in the process of analysing big data are heterogeneity and incompleteness of information, scalability, privacy of information, timeliness and lack of available analytical skills (Agrawal et al., 2011; Cuzzocrea et al., 2011; Mohanty et al., 2013). Also cost, technology.

### ***Heterogeneity and incompleteness***

Analytical algorithms are not designed to work with heterogeneous data (Agrawal et al., 2011). Data sources often store heterogeneous data and before analysis can take place the data must be cleaned and transformed into a structured format in order to perform analysis (Cuzzocrea et al., 2011). Even after cleaning and transformation of the data, however, there may still be errors and incompleteness of information which needs to be managed during data analysis (Agrawal et al., 2011).

### ***Scalability of analytic algorithms***

As big data is concerned with high volumes, one of the challenges to data analysis is the scalability of the algorithms which are used for analysing big data (Chen and Zhang, 2014; Rajakumar and Sushma, 2013). The scalability of the algorithm is the ability of the algorithm to scale rapidly with increasing dataset volumes (Kaisler et al., 2013). As the size of data grows, the time taken to access that data becomes less and less efficient (Jacobs, 2009). Analytical algorithms should therefore be developed to ensure scalability as datasets grow in volume (Kaisler et al., 2013; Rajakumar and Sushma, 2013). The challenge for retailers was in selecting analytical algorithms which would allow them to cope with large volumes of data (Chen and Zhang, 2014; Kaisler et al., 2013).

### ***Privacy of information***

Preserving individual privacy of information is a challenge when analysing big data (Michael and Miller, 2013; Mohanty et al., 2013). Retailers can capture information about an individual from a variety of sources such as social media and loyalty programs (Manyika et al., 2011). As big data platforms can aggregate information on an individual, the challenge for retailers will be in ensuring that customer information is not divulged (Sawant and Shah, 2013).

### ***Timeliness***

Another challenge of big data analysis is analysing data in a timely manner (Mohanty et al., 2013). The larger the dataset that is being analysed the longer the time required

for analysing the dataset (Mohanty et al., 2013). The time span of the analysis is important because it affects how quickly decisions can be made in response to a change in the business environment (Ularu et al., 2012). Retailers have to ensure that the size of the dataset they are working with allows them to analyse data timeously (Ularu et al., 2012).

### ***Availability of analytical skills***

In order to capitalise on the potential of big data there is a need for people with the skills necessary to make discoveries from big data (Davenport & Patil, 2012; Manyika et al., 2011). Data scientists and other professionals possess the technical skills required for working with and analysing big datasets but are scarce and in high demand (Manyika et al., 2011; McAfee & Brynjolfsson, 2012). The challenge for retail companies is in identifying and attracting people with the necessary skills to the enterprise in order to take advantage of big datasets (Davenport and Patil, 2012).

### **Techniques and technologies for big data analytics**

Techniques and technologies have been developed which can help capture, store and analyse big data to obtain information which is of value for decision making (Chen and Zhang, 2014; Manyika et al., 2011). An overview is given on some of the main techniques and technologies which can make up a big data analytics solution.

#### ***Techniques for big data analytics***

There are a wide variety of techniques available which can be employed to derive insights from big data and many of these techniques draw from existing disciplines such as mathematics and statistics (Manyika et al., 2011). An overview of some of the techniques is given.

#### ***Optimisation methods***

Optimisation methods use a set of numerical techniques to redesign a system or process in order to improve performance according to a certain measure (e.g.: speed and cost) (Manyika et al., 2011). Optimisation methods can be applied to improve retailers' operational processes such as inventory management, price optimisation floor layout and product range strategies (Manyika et al., 2011).

#### ***Data mining***

Data mining is a set of techniques, such as association rule learning, cluster analysis, classification and regres-

sion, which are used to extract patterns from data (Chen and Zhang, 2014; Manyika et al., 2011). Retailers use data mining for: determining which segments are more likely to respond to an offer, identifying characteristics of successful employees and performing a basket analysis to determine customer purchase behaviour (Manyika et al., 2011). Big data mining is more challenging than more traditional data mining algorithms because existing techniques have to be extended in order to deal with the larger workload (Chen and Zhang, 2014).

#### ***Neural networks***

Neural networks are computational models which are based on biological neural networks and are used for detecting patterns in data (Manyika et al., 2011). Neural networks can be used for pattern recognition, image analysis, optimisation and adaptive control (Chen and Zhang, 2014; Manyika et al., 2011). Retailers apply neural networks to: identify high-value customers which are at risk of leaving a particular company and detect fraudulent insurance claims (Manyika et al., 2011).

#### ***Machine Learning***

Machine learning is a part of the field of artificial intelligence and involves the design of algorithms which allow computers to adapt behaviour based on empirical data (Chen and Zhang, 2014; Manyika et al., 2011). An important focus of machine learning is on discovering information by detecting patterns and making intelligent decisions based on information (Chen and Zhang, 2014; Manyika et al., 2011). Machine learning can help retailers to become more precise and granular in making predictions (Davenport, 2013). Examples of machine learning include natural language processing, association rule learning and ensemble learning (Chen and Zhang, 2014; Manyika et al., 2011).

#### ***Cluster analysis***

Cluster analysis uses techniques to break down a diverse group into smaller groups of object with similar characteristics (Manyika et al., 2011). Retailers use cluster analysis for segmenting consumers into groups in order to perform targeted marketing (Manyika et al., 2011).

#### ***Predictive modeling***

Predictive modelling uses a set of models in order to predict the probability of an event occurring (Manyika et al., 2011). Predictive modelling can be applied in the retail industry to predict the churn rate of customers or the likelihood that a customer can be cross sold another

product (Manyika et al., 2011).

### **Technologies for big data analytics**

There are a growing number of technologies which can be used for analysing and managing big data (Manyika et al., 2011). Some of the major technologies which support techniques for analysing big data are discussed.

#### ***Data warehouse***

A data warehouse is a database that is used for storing copies of transaction data which are specifically structured for query and analysis (Chen and Zhang, 2014; Jacobs, 2009). Data warehouses and data marts (a subset of a data warehouse) are commonly used for managing the storage, retrieval and analysis of structured big datasets (Chen and Zhang, 2014). Data warehouses utilise extraction, transformation and loading processes to transform big datasets into a structured format for storage and retrieval (Manyika et al., 2011).

#### ***Distributed systems***

A distributed system consists of multiple computers connected together through a network which are used to solve a common computational problem (Manyika et al., 2011). The problem to be solved is broken down into tasks which are solved by one or more computers working together in parallel (Manyika et al., 2011). Distributed systems are useful for analysing big datasets (Jacobs, 2009).

#### ***Extraction, transformation and loading (ETL)***

ETL tools are designed to extract raw data from a variety of sources, transform the data according to a predefined structure and load the data into a database or data warehouse (Cuzzocrea et al., 2011; Manyika et al., 2011). In order for big data to be analysed effectively big data sources have to be transformed into a structured format for storage (Cuzzocrea et al., 2011). ETL processes help transform big data sources into a suitable structured format so that data can be analysed to obtain meaningful information (Cuzzocrea et al., 2011).

#### ***Hadoop***

Apache Hadoop is an open source software framework for writing applications which process large datasets in parallel on a distributed system (Chen and Zhang, 2014; Fan and Bifet, 2013; Manyika et al., 2011). Hadoop allows

organisations to load, store and query large datasets on multiple servers and perform advanced analytics in parallel (Davenport et al., 2012).

### **Vendor products for big data analytics**

There are a number of vendors which provide tools and platforms for big data analytics (Singh and Singh, 2012). The following vendor products are the key big data platforms being offered and each product represents a different approach to big data analytics (Russom, 2011; Singh and Singh, 2012).

BigInsights is a product from IBM which is based on the open source Hadoop framework and IBM's Big Sheet text analytics module (Russom, 2011; Singh and Singh, 2012). BigInsights is designed to help manage and analyse high volumes of structured and unstructured data using features such as text analytics for data discovery and exploration (Russom, 2011; Singh and Singh, 2012). IBM InfoSphere Streams is a platform offered by IBM which allows streams of data to be analysed and processed in real time (Russom, 2011; Singh and Singh, 2012). IBM InfoSphere Streams is a scalable and agile platform which allows big data analytics on a variety of structured and unstructured data types (Russom, 2011; Singh and Singh, 2012).

SAP HANA is an in-memory appliance, released by SAP, which allows analytical queries to be run in real time against detailed datasets without having to transform the data into a structured format for analysis (Russom, 2011). HANA implements a variant of the MapReduce to allow queries to be run without the need for a data model (Russom, 2011).

EMC Greenplum Database is a massively parallel processing (MPP) architecture (Russom, 2011) which allows for low latency access to large volumes of data (Mohanty et al., 2013).

The SAND Analytic Platform is a columnar analytic database platform which allows for data scalability by using MPP (Russom, 2011; Singh and Singh, 2012). The SAND platform is designed to support thousands of users concurrently and provide features such as query optimisation, in-memory analytics and full text search (Russom, 2011; Singh and Singh, 2012). The focus of the SAND analytic platform is analysing complex tasks such as customer loyalty programs and customer churn rates (Russom, 2011; Singh and Singh, 2012).

### **Summary**

From the literature review it is apparent that big data and big data analytics are not clearly defined terms. Big data analytics can be applied in the retail sector to assist with analysing unstructured data through customer sentiment analysis, optimising prices and managing inventory. There

are, however, barriers to using big data analytics which need to be overcome, such as the privacy of information and scalability of analytic algorithms. In order to deal with the complexity of big data, retailers can use various analytic techniques and technologies to help analyse big data. There are also a variety of off-the-shelf products available, such as IBM BigInsights and SAP HANA, for managing and analysing big data to help with supporting decision making.

The purpose of this study is to assess the use of big data analytics in the retail industry in South Africa, and from the literature review the following research questions emerged:

- RQ1 How are South African retail companies defining big data and big data analytics?
- RQ2 To what extent are South African retail companies using big data analytics?
- RQ3 What value can South African retail companies gain from using big data analytics?
- RQ4 What are the barriers for South African retail companies in implementing big data analytics?
- RQ5 What techniques and technologies are South African retail companies using for analytics and big data?
- RQ6 What vendor products are available to South African retail companies?

## RESEARCH METHODOLOGY

The research philosophy adopted reflects important assumptions about the researchers' view of the world and what constitutes "valid research" (Myers & Avison, 2002). These assumptions are important because they underpin the research strategy and they help to clarify what is being investigated (Saunders et al., 2009). The underlying ontological assumption of the study was interpretive because it was believed that reality is constructed as a result of the actions and perceptions of social actors (Bhattacharjee, 2012; Saunders et al., 2009). Thus it was believed that the best way to assess the usage of big data analytics was through the subjective interpretation of the respondents involved (Bhattacharjee, 2012).

### Sample and data collection

The majority of retailers in South Africa are small (Charman et al., 2015), and would be unlikely to be using big data. The target population for the study was medium to large South Africa retailers, big data analytics vendors, and professional service companies in order to get three different perspectives on the usage of big data analytics in South African retail companies. The study purposively targeted retailers in Johannesburg and Cape Town who used business intelligence from various sub sectors such as fashion, food and pharmaceuticals. IT professionals and vendors, with experience in the field of business intelligence were selected based on recommendations from some of the retailers interviewed (see Appendix A for interview questions). A qualitative research approach was employed because it allowed for individual views and opinions to be explored. Although the questions were structured the interview process was conducted in such a way that follow up questions could be asked. Once respondents agreed to take part and a date was confirmed, the interview questions were emailed ahead of the interview to ensure respondents had read and understood the questions. In total, 13 interviews were conducted

over a period of three weeks. Nine of the 13 interviews were face to face and conducted in Cape Town. Two telephone interviews and two Google Hangouts interviews were conducted for convenience as some of the respondents were based in Johannesburg. Interviews lasted between 25 minutes to an hour and 17 minutes. All interviews were recorded and were transcribed. The interviews questions were adapted from a study published by The Data Warehousing Institute (Russom, 2011) on big data analytics and were arranged according to the major themes that were identified in the literature review. Open ended questions were also included in order to probe answers from respondents.

### Data analysis method

Thematic analysis was selected as the data analysis method for this study, as it was appropriate for determining patterns and themes relating to the use of big data analytics (Braun and Clarke, 2006).

Patterns and themes were identified by following the process of performing thematic analysis down into six phases as briefly discussed below.

1. Becoming familiar with the data that had been gathered from the semi structured interviews. The key approach to this was in the transcription and comparison of this to the original interviews for accuracy.
2. Reading through all of the transcriptions and generating codes which described interesting features of the responses from respondents. Atlas ti was used to help assist with the coding process. The creation of codes involved highlighting text or lines in the transcription and tagging and naming the selected text.
3. Searching for themes involved reading through all of the codes and assigning them a common theme. Examples of initial themes were: usage, challenge, definition, barriers to adoption, future, architecture, value, challenge, technologies, maturity, perception and motivation.
4. Organising the codes under their respective themes. Themes such as "Maturity" and "Concerns" were removed because there was insufficient evidence to justify the theme.
5. Rereading the coded extracts under each theme in order to identify subthemes. The process of identifying subthemes helped to group common codes together and to remove some of the codes which did not form part of a logical grouping. Themes were also given final names and subthemes in preparation for the final write up. Writing up the themes which had been identified as set out below.

## DATA ANALYSIS AND FINDINGS

The purpose of the research was to assess the use of big data analytics in the retail industry in South Africa. In order to achieve the aim of the research, information was collected by interviewing retailers, big data analytics vendors and professional services. The data collected from respondents was analysed using thematic analysis and the findings have been structured around the core themes which were identified in the literature review.

### Description of sample

Respondents were grouped into three broad sections, Retailers, Vendors and Professionals. Respondents were given pseudonyms based on the section they came from, all Retailers were given pseudonyms starting with an R,

**Table 1.** Retailers interviewed.

Interview	Pseudonym	Role within organisation	Years in IT	Organisation	Number of employees in organisation
1	Rob	General Manager for BI	20+	Retailer 1	38 000+
2	Rachel	BI manager	20+	Retailer 2	23 000+
3	Rick	Business process owner for BI	12+	Retailer 1	38 000+
4	Rebecca	BI manager	25+	Retailer 3	8400+
5	Robin	Head of merchandise systems	12+	Retailer 4	17 000+
6	Russell	Head of BI	20+	Retailer 5	9700+
6	Roy	BI Developer	3+	Retailer 5	9700+
6	Ross	Software Developer	8+	Retailer 5	9700+
7	Rodney	BI analyst	9+	Retailer 6	4000+
8	Rupert	Customer data and insights	20+	Retailer 7	100 000+

**Table 2.** Big data vendors interviewed.

Interview w	Pseudonym	Role within organisation	Years in IT	No of employees in organisation	Industry
9	Vincent	Retail architect	18+	430 000+	Computer hardware, software and IT consulting
10	Victor	Presales expert	18+	66500+	Enterprise software
11	Vernon	Strategic account director		120 000+	Enterprise software
11	Vaughan	Presales consultant	20+	120 000+	Enterprise software

**Table 3.** Professional services interviewed.

Interview w	Pseudonym	Role in organisation	Years in IT	Industry	No of employees in organisation
12	Peter	Managing Director	14+	BI solutions and BI consulting	350+
13	Patrick	Strategic consultant and head of BI	20+	Professional services, BI and software development	51+

Vendors were given pseudonyms starting with a V, and Professionals were given pseudonyms starting with P. Details of the respondents are set out in Tables 1 to 3.

### The state of big data analytics in the retail industry

The majority of the retailers interviewed were not generating or analysing big data, because they could not find a use case for big data. The big data vendors and professional services confirmed that retailers were not using big data but that some of them were in the process of investigating it (Rachel, Rebecca, Robin, Russell, Rupert, Peter, Patrick, Victor, Vincent).

*“[you might] presuppose that large retailers are making use of big data in a constructive way...and I think you will find that disappointing that we aren't...”* (Russell).

*“Retailers are not using big data in SA... I think one of the challenges in the retail world is there are*

*people who are struggling to understand what real tangible business value big data is.”* (Patrick).

*“There are retailers that are local that are looking to do this...both your consumer goods type as well as the more fast moving consumer goods and the pure retail...they are exploring how they can go about doing this while they are still trying to get their backend sorted out...”* (Victor).

*“From a retailer perspective I think some of them are semi doing it or they are busy in the adoption process ...”* (Peter).

Two retailers mentioned using big data analysis in a limited way:

*“We are certainly using it for sentiment analysis...so some of the analytics is hired out to third party vendors who specialise in things like sentiment analysis and surfing your social media and understanding what the different comments are...”* (Robin).



*"Yes we are...so we have got an area where we actually employ data scientists... and we employ a tool that is designed for analytics being SAS" (Robin).*

### Defining big data

The interviews reflected a clear lack of consensus on the definition of big data. Some used the term big data interchangeably with business intelligence:

*"We have always called it BI ... Those things ... have been done all along... it's not as if those things made big data..." (Victor)*

*"BI or big data in a retail environment is fundamental..." (Vernon).*

Others viewed big data as something different but were somewhat unclear about this difference. For example: *"if your current data warehouse cannot deal with the data that you need to answer specific business questions, you probably have a big data problem" (Rachel)*, and as all the information that is critical to the operating of an organisation (Robin).

*It is just large amounts of variety and volumes of data... and structured and unstructured data..." (Peter).*

Some expressed a cynical view of big data and said that the term was created in order to sell big data platforms (Rob, Russell)

*"...in reality the vendors ... are running around and looking for some new way to sell their software and their ideas and so now it is big data" (Rob)*

*"...on one side you can be saying this is just hype and a marketing way of selling hardware because it is nothing new like cloud computing and mobile..." (Russell)*

There was some recognition of the three Vs or attributes of big data (Russell, Vincent). When prompted, there were views expressed on these attributes.

### Volume

The majority of respondents viewed the volume of data in the organisation in terms of terabytes of data stored on physical disks. Respondents also quantified the volume of data in terms of: number of POS transactions, millions of transactions, number of rows in a table and number of stock keeping units (SKU). These measures appeared to vary in different parts of the organisation.

*I actually go by the TB and so many years' worth of sales history..." (Rebecca)*

*I think certainly in terms of an IT perspective we will*

*work in size and number, so how many terabytes is this system and how many terabytes is that system. From a BI perspective we regularly look at number of rows..." (Rick)*

*I would measure how many stores I have and how many articles do I have in a range or average per store...that will give you the amount of SKUs you are dealing with..." (Rob)*

### Velocity

Respondents suggested that retailers in South Africa mainly process their data in batch processes but aim to process data in real time (Rob, Rachel, Peter, Rupert).

*"... traditionally it has been batch or reactive type analysis. The data probably gets inserted in real time or near real time about 10 - 15 minutes..." (Peter)*

*"I can't think of an example where we're doing real time monitoring and processing of data. Although it's obviously something to aim for" (Rupert).*

### Variety

Retailers in South Africa are collecting structured and unstructured information but the majority of data is structured transactional data (customer sales). The variety of information is therefore small for retailers in South Africa. In addition, retailers appear to be focused on deriving value from their structured data where they can see tangible business value as opposed to their unstructured data (Vincent, Rick).

*"I would even go as far as saying 90% are structured. You are only just starting to see trends where retailers are starting to tap into unstructured data" (Vaughan).*

Rupert was the only respondent that mentioned tapping into Unstructured Supplementary Service Data (USSD),

*"...we have POS data which is one of the main sources of data... We don't have a loyalty programme, but we have various loyalty mechanisms ...We use USSD extensively..."*

### Value

Value was one area where the respondents were able to express opinions quite freely even though they were negative about value. Retailers in South Africa appear to be focused on deriving value from their structured data where they can derive tangible value as opposed to their unstructured data which they are struggling to find a use case for it.

*"We don't generate data that we can't deal with in our current environment, so it's actually quite hard to find a use case for big data in retail."* (Rachel).

*"We are going to start using it, but it's not going to be a whole big bang thing. So we're going to do it use case by use case and business case by business case ..."* (Rebecca)

Some retailers confirmed that they would potentially start using big data analytics once they had perfected the mining of their structured transactional data (Rachel, Rick, Rob).

Retailers cannot seem to find a use case for big data but they admit that there is possible value in tapping into big data in the future (Rachel, Russell).

*"It's definitely something that we, as part of our BI strategy...have this big data strategy that we are looking at"* (Rachel).

There was limited potential value which the respondents thought retailers could obtain from big data analytics. Big data analytics could be used to analyse customer information in order to market to customers on an individual basis (Peter, Rebecca, Vaughan).

*"...any of these companies need to know who their customer is... knowing that they can start really target marketing towards an individual...So big data has a lot of value in that sense"* (Peter).

*"... the more you can find out about a customer, the better you can promote to them. Ok, that would be a good use case"* (Rebecca).

Big data analytics can be used to help with merchandising or inventory management (Patrick, Rupert).

*"Merchandising for me is also another big area because it is looking at sales patterns based on a whole bunch of micro and macro-economic factors"* (Patrick)

*"...well it is three areas: one is merchandising-just set the products out in the store, other is price, and the other is promotion..."* (Rupert).

### **Defining big data analytics**

Most of the respondents were unclear about big data analytics or devised their own definitions (Victor, Robin, Rodney, Peter). Rob believed the term big data was nothing new and was invented to help sell software,

*"I just think it is a hype and a name ... When you get down to brass tacks there is only certain types of data and they are in certain volumes and areas."*

There were respondents who perceived big data analy-

tics to be a repackaging of existing data analysis tools but somehow bigger and faster (Rachel, Rebecca, Rick, Victor).

*"there's very, very little revolutionary programming being done... if anything it's a case of throwing hardware, of throwing hardware at all of this data to try and make some sense of it"* (Rodney).

### **Usage of big data analytic platforms**

While some were not using big data analytics they admitted to using big data analytics platforms for processing information more quickly in parallel (Patrick, Rachel, Rob, Rupert)

*"I happen to know that those that are saying they are doing big data are actually not doing big data. It's really just they are taking big data platforms..."* (Patrick).

However, the majority of the retailers interviewed were not using sentiment analysis for their social media data as they had not yet found a use case to justify the investment in sentiment analysis.

There was some suggestion of a use of big data analytics to assist with managing the risk of granting credit in the organisation but this was not really big data analytics.

*"So one of the things that was construed as big data is when we do a credit check...so we want to sell merchandise to a potential customer... and we do a credit check to find out if this person pays their account and we grant them some credit..."* (Robin).

*"There are guys out there looking at predictive analytics around how to manage your credit risk. I think there are lots of opportunities there with big data and maybe machine learning to start to come up with new insights and new ideas around managing credit risk ..."* (Patrick).

Big data analytics can be used to better understand the customer and to service them better by providing a personal experience (Rupert, Victor).

*"Customer's user experience and buying behaviour is the most important thing from any data analysis point of view in a consumer facing organisation..."* (Rupert).

*"I think is the big value proposition for most of the retailers."* (Victor)

### **Barriers to using big data analytics in retail**

The most likely barriers for retailers are: availability of analytical skills, costs of big data analytics platforms, contextualising big data to provide meaning, return on

investment of big data ventures, leadership buy in, product selection and the need to structure current data.

There was a general recognition by respondents of the importance of having skilled individuals to discover insights from big data (Rachel, Rebecca, Roy, Rob, Vaughan).

*"I think the biggest problem if you look at the SA space is the skills of the people who are capable of doing this stuff" (Roy).*

*"...you also have to have skilled individuals in terms of structuring the data and putting it in context and visualising it...skills are resources which are a challenge..." (Rob).*

*"...there would be challenges in terms of skills in terms of transforming the concept of big data into reality...it's about integrating all of these other types of data that we don't understand..." (Vaughan).*

*"They are very scarce skills so, you are going to pay for it if you outsource it, but if you up skill your own team then that means you take people out of project teams or out of operational teams. ... [also] it's a complete mind shift for people who have worked in traditional BI for a while" (Rachel).*

*"They all work differently, there's no expertise readily available" (Rebecca).*

A significant barrier to using big data analytics is that South African retailers are primarily focused on mining structured data whereas big data analytics by definition involves the analysis of different types of data (Rob, Rick, Russell, Patrick).

*"We have got a pipeline of plenty stuff that is going to keep us going for ten years just on the structured stuff" (Russell).*

One of the major barriers is the cost of the big data analytic platforms (Rachel, Patrick, Rob). The data suggested that big data analytic platforms are costly investments.

*"No it's millions. It's absolutely millions" (Rachel).*

Another potential barrier for retailers is contextualising big data from the context of the organisation in order to make business decisions (Peter, Rebecca, Rick).

*"Because from my point of view you need to have some context for the data to actually be able to use it at all. The big thing I think with big data is you have to be able to contextualise it somehow." (Rebecca)*

*"People don't intuitively know how to contextualize information in the context of their organization..." (Rick).*

### **Buy in from leadership**

As mentioned above, retailers are struggling to find a use

case to justify investing in big data. Not only did the executive lack understanding of big data but, they felt that the return on investment does not justify investment (Patrick, Rick, Rob, Rodney, Rupert, Russell).

*"I think one of the challenges in the retail world is there are people who are struggling to understand what real tangible business value big data is...." (Patrick).*

*"... they might have a big data strategy but the value that it is going to add to the business is decimals of a percentage in terms of the return on investment..." (Rob).*

*"...well it's no different from the value you can get from analysing data that's stored in traditional data warehousing environments..." (Rupert).*

*"...if you don't have the support of the business in this, I'm not quite sure why you would even embark on it." (Rachel).*

### **Product selection**

The choice of a big data analytic platform was also a barrier because of the number of platforms available (Rachel, Patrick, Victor).

*"You have to decide which of those you want to go with." (Rachel).*

*"For me that's also potentially one of the barriers at the moment because there is so much out there at the moment." (Patrick).*

### **Need to structure big data**

Another barrier that was identified is the need to structure the various data types so that information can be understood and analysed. Respondents identified that structuring data in a format in which it can be consumed is a challenge in general (Rachel, Rob, Rick, Rupert).

Rob said that, *"The problem in most of the retail world is that people haven't structured their structured data properly ... and put it in context so it is useable..." (Rob).*

### **Techniques and technologies for big data analytics**

While the majority of retailers were not using big data analytics they mentioned using analytical techniques such as data mining, exception reporting, predictive analytics and statistical analysis. Of these, the most frequently mentioned were data mining (Peter, Russell) and exception reporting (Rob, Rupert, Vincent). Data mining was used for relatively ordinary applications such as monitoring refrigeration (Peter) and improving marketing campaigns and for predicting the number of staff members they required for a particular outlet (Russell). There were few examples of exception reporting other

than to highlight changes in sales trends (Vincent).

## Technologies

A number of data analysis technologies were discussed. Many of these could not be construed as big data analysis technologies. Several respondents mentioned using data warehouses from vendors such as Oracle, Microsoft and SAP for storing big datasets in a structured format for storage and retrieval (Rebecca, Rick, Rodney, Russell, Victor).

*"The biggest challenge is the fragmentation of information around the business. Most retailers have probably 3-5 data warehouses serving different purposes..."* (Victor).

It appears that no retailers have ventured into Hadoop but some are currently investigating using Hadoop (Peter, Patrick, Rebecca, Rachel, Rupert).

A few mentioned using predictive analytics tools such as SAS and Cognos for predicting customer behaviour, risk and credit (Robin, Russell, Victor, Vincent).

Interestingly, three retailers mentioned using Microsoft Excel for analytics (Rick, Robin, Russell).

*"The one piece of technology that keeps most organisations alive is Microsoft Excel...it's one of the most underrated taken for granted application and yet it is used for the most amazing things"* (Robin).

A number of respondents mentioned using software such as Tableau and SAP business objects as a front end for data visualisation (Rick, Rebecca, Rob, Robin, Russell).

*"...people should be spending more time and effort to assist in visualisation...so we go to companies like Tableau..."* (Rob).

Some of the vendors mentioned offering a SQL environment which removed the barrier of writing MapReduce to query the Hadoop environment (Vaughn, Vincent) but it appears that this has not reached the retailers. Victor and Vaughan mentioned offering complex event processors as technologies to help analysing information as it travels in flight in real time. No retailers, however, mentioned using complex event processors.

## Vendor products for big data analytics in South Africa

Respondents reported a number of products being offered by vendors such as IBM, SAP, EMC Greenplum, Cloudera, Oracle and SAS which are available to retailers in South Africa. There has been limited take up of such products in the retail companies surveyed. In fact there is

*"Not that much from a big data analytics perspective"* (Vincent).

Rob considered a number of in-memory platforms such as Hana but decided to go with Greenplum, as did Rupert, while Roy had chosen SAP Hana. Rachel was considering Greenplum but was not going to go with it because it was too costly. Rachel clearly expressed the current situation:

*"Actually using it for big data and big analytics - implementing it, so it's established and operational...I'm not sure any South African company has done it successfully."* and went on to say: *"if we do big analytics, it would be SAS"*

## DISCUSSION

The purpose of the study was to assess the use of big data analytics in the retail industry in South Africa. The usage of big data analytics was assessed by incorporating perspectives from retailers, vendors and professional services in order to obtain a broader interpretation of the use of big data analytics in retail.

A dominant theme in the findings was a lack of consistent and clear definition of big data and big data analytics and this is shown by the multitude of definitions used by industry professionals. This also reflected a general lack of understanding of big data analytics in the organizations surveyed. However, most either remembered or picked up on a mention of the defining attributes of Chen and Zhang (2014), of volume, velocity, variety and value. The respondents were able to discuss volume both in terms of terabytes of data stored and other measures such as number of POS transactions, millions of transactions, number of rows in a table and number of stock keeping units (SKU). As most processing is batch, there was little comment on velocity.

The majority of data in the surveyed organisations is structured transactional data (customer sales) resulting in little variety of data. The focus is still on structured data where the tangible business value can be seen. Some sources also indicate that there might be a fourth V which represents the value that could be created by using big data to enable enhanced decision making, to get insight discovery, and to optimize processes (De Vries, 2013b). However, the fourth id not generally accepted. The respondents, however, did express views on value.

Most respondents perceived big data analytics to be nothing new, whereas others viewed big data analytics as an evolution of current technology. A clear understanding of big data and big data analytics would be beneficial, as a retailers' perception of big data analytics may influence their reaction to the concept. Many respondents indicated that big data is not a new concept, a viewpoint which is evident in the literature (Russom, 2011). However, it is important to note that, while big data may not be a new

concept, the processes and technologies involved in analysing big data are new and should be treated as such.

The main finding of the study was that retailers were not using big data and therefore were not performing big data analytics. Robin was the only respondent who mentioned using big data analytics. South African retailers focus primarily on the analysis of structured data, whereas big data analytics by definition, involves the analysis of different data types. The findings suggest that retailers have yet to find significant business value in analysing unstructured and semi-structured data to justify the investment of doing big data analytics. As a result their unstructured data is underutilised. Customer sentiment analysis, for example, which analyses customer social media data, is not commonly performed amongst retailers because the return on investment is difficult to justify.

While South African retailers are not analysing big data, there is a drive for the use of big data analytic platforms to analyse structured data. Retailers recognise the potential of using big data analytics platforms for solving issues in their current Business intelligence environment. Few were in a position to even comment on the use of big data analytics for price optimization, customer micro-segmentation and targeting, sentiment analysis and in-store behaviour analysis. There was some mention of inventory management but without the use of complex forecasting

The South African retailers interviewed recognise that additional value can be gained from using big data analytics to tap into unstructured and semi-structured data. However, they believe that big data analytics should be done once they have become mature in terms of analysing their structured data.

Before South African retailers can start using big data analytics there are significant barriers that need to be overcome. A substantial capital investment is required in order to purchase a big data analytic platform and the investment has to be justified sufficiently in order to obtain leadership buy-in, few saw the possibility of making a business case for this. The lack of available of analytical skills is also a barrier to using big data analytics. Retailers may have to either up-skill their employees or outsource in order to deal with the analysis of different data types.

Barriers discussed in the literature review such as heterogeneity and incompleteness, scalability of analytical algorithms, privacy of information and timeliness were not even contemplated by the respondents.

While retailers are not using big data analytics, there are a number of techniques and technologies which are being used for traditional analytics. Certain techniques (data mining) and technologies (data warehouses) which retailers mentioned using aligned with the findings in the literature review. Retailers are therefore using techniques

and technologies which can be applied to big data for performing traditional analytics. Technologies that were not mentioned in the literature review but mentioned include Microsoft Excel and complex event processing. Complex event processing may become of particular interest to retailers in the future because it provides the ability to analyse data in real time. While retailers mainly process their data in batches, there is potential to use complex event processing once the need for real time analytics arises. In summary, the research has shown that big data analytics requires a significant investment by the retailer, which many retailers cannot currently justify. This represents the main limitation to using big data analytics in the South African retail industry, and while future use-cases may be found, the value of big data analytic platforms remains in their ability to rapidly analyse existing, structured data sets.

## Conclusion

The purpose of this study was to assess the use of big data analytics in the retail industry in South Africa. To this end, big data analytics usage in South Africa was assessed by investigating usage-related factors, such as big data analytics products, techniques and technologies. The findings showed that retailers are not using big data analytics because there is a focus on exploiting existing structured data completely before tapping into unstructured and semi-structured data. Some retailers are, however, leveraging the enhanced processing speeds of big data analytic products to improve on traditional analytics.

While the findings indicate that retailers cannot find a worthwhile use case for big data analytics, there may however be potential for assessing the usage of big data analytics in other industries. For instance, there may perhaps be a need for big data analytics in the telecommunications industry where there are large volumes of structured call data.

From the findings, it is evident that retailers in South Africa are not using big data analytics, due to the lack of an obvious use-case to justify the implementation costs. Furthermore, the study has shown a debate around the definition of big data and big data analytics, as well as a multitude of conflicting perceptions on the topic. Future research into big data analytics should consider how these perceptions influence decision making, and how this affects the future usage of big data analytics in the retail industry.

## ACKNOWLEDGEMENT

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number 91022).

## Conflict of Interests

The authors have not declared any conflict of interests.

## REFERENCES

- Agrawal D, Bernstein P, Bertino E (2011). Challenges and Opportunities with big data 2011-1. Proceedings of the VLDB Endowment (pp. 1–16). West Lafayette. Retrieved from <http://dl.acm.org/citation.cfm?id=2367572>
- Bhattacharjee A (2012). Social Science Research: principles, methods, and practices. Book 3. (second., pp. 1–147). Open Access Textbooks. Book 3.
- Braun V, Clarke V (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Brown I, Russell J (2007). Radio frequency identification technology: An exploratory study on adoption in the South African retail sector. *Inter. J. Informat. Manage.* 27(4), 250–265. doi:10.1016/j.ijinfomgt.2007.02.007
- Camm J, Cochran J, Fry M, Ohlmann J, Anderson D (2014). A Categorization of Analytical Methods and Models. In *Essentials of Business Analytics* (pp. 5–7). Cengage Learning.
- Charman AJ, Petersen LM, Piper LE, Liedeman R, Legg T (2015). Small area census approach to measure the township informal economy in South Africa. *J. Mixed Methods Res.* P.1-23.
- Chen CL, Zhang CY (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, Upcoming, 1–34. doi:10.1016/j.ins.2014.01.015
- Chen H, Chiang R, Storey V (2012). Business Intelligence and analytics: From big data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
- Cooper A (2012). What is Analytics? Definition and Essential Characteristics. *CETIS Analytics Series*, 1(5), 1–10.
- Cuzzocrea A, Song I, Davis K (2011). Analytics over large-scale multidimensional data: the big data revolution! In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP* (pp. 101–103). New York: ACM.
- Davenport T (2013). Analytics 3.0. *Harvard Busin. Rev.* 91(12), 64–72.
- Davenport T, Barth P, Bean R (2012). How “big data” is different. *MIT Sloan Management Review*, 54(1), 22–24.
- Davenport T, Dyche J (2013). Big data in Big Companies (White paper). May 2013 (pp. 1–31). USA. Retrieved from <http://swqlab.fov.uni-mb.si/moodle/pluginfile.php/9774/course/section/534/Big-Data-in-Big-Companies.pdf>
- Davenport T, Patil D (2012). Data Scientist: The sexiest job of the 21 st century. *Harvard Bus. Rev.* 90(10), 70–76.
- Delen D, Demirkan H (2013). Data, information and analytics as services. *Decision Support Systems*, 55(1), 359–363. doi:10.1016/j.dss.2012.05.044
- Devlin B, Rogers S, Myers J (2012). Big data comes of age (whitepaper) (pp. 1–43). Retrieved from [http://www-03.ibm.com/systems/hu/resources/big\\_data\\_comes\\_of\\_age.pdf](http://www-03.ibm.com/systems/hu/resources/big_data_comes_of_age.pdf)
- Fan W, Bifet A (2013). Mining big data: current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter*, 14(2), 1–5. Retrieved from <http://dl.acm.org/citation.cfm?id=2481246>
- Fisher D, DeLine R, Czerwinski M, Drucker S (2012). Interactions with big data analytics. *Interactions*, 19(3), 50–59.
- Jacobs A (2009). The pathologies of big data. *Communications of the ACM*, 52(8), 36–44.
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and Challenges Moving Forward. In *2013 46th Hawaii International Conference on System Sciences* (pp. 995–1004). Maui: IEEE. doi:10.1109/HICSS.2013.645
- Kwon O, Lee N, Shin B (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *Inter. J. Informat. Manage.* 34(3):387–394. doi:10.1016/j.ijinfomgt.2014.02.002
- Madsen M (2013). The Challenges of big data & Approaches to Data Quality (White paper) (pp. 1–10). Rogue River. Retrieved from [http://fm.sap.com/data/UPLOAD/files/The\\_Challenges\\_of\\_Big\\_Data\\_&\\_Approaches\\_to\\_Data\\_Quality.pdf](http://fm.sap.com/data/UPLOAD/files/The_Challenges_of_Big_Data_&_Approaches_to_Data_Quality.pdf)
- Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers A (2011). Big data: The next frontier for innovation, competition, and productivity (pp. 1–143). Retrieved from <http://www.citeulike.org/group/18242/article/9341321>
- McAfee A, Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 1–9.
- Michael K, Miller K (2013). Big data: New opportunities and new challenges. *Computer*, 46(6), 22–24.
- Mohamed S, Ismail O, Hogan O (2012). Data equity Unlocking the value of big data (Whitepaper) (pp. 1–44). London.
- Mohanty S, Jagadeesh M, Srivatsa H (2013). Big data Imperatives: Enterprise “big data” Warehouse, BI Implementations and Analytics (pp. 1–296). California: Apress.
- Myers M, Avison D (2002). An introduction to qualitative research in Information Systems. In *Qualitative Research in Information Systems* (pp. 3–12). London: Sage Publications.
- Rajakumar M, Sushma M (2013). Challenges and Opportunities of big data Analytics in Business Applications. *Inter. J. Electro. Communicat. Computer Engineer.* 4(6), 6–9.
- Retail trade sales. (2013) (pp. 1–14). Retrieved from <http://www.statssa.gov.za/publications/P62421/P62421March2013.pdf>
- Russom P (2011). Big data analytics. TDWI Best Practices Report, Fourth Quarter (pp. 1–35). Renton. Retrieved from <http://public.dhe.ibm.com/common/ssi/ecm/en/iml14293usen/IML14293USEN.PDF>
- Saunders M, Lewis P, Thornhill A (2009). Research methods for business students (fifth., pp. 2–614). London: Pearson Education Limited.
- Sawant N, Shah H (2013). Big data Application Architecture. *Big data Application Architecture Q & A* (pp. 1–143). California: Apress.
- Shmueli G, Koppius O (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, 35(3), 553–572.
- Singh S, Singh N (2012). Big data analytics. In *2012 International Conference on Communication, Information & Computing Technology (ICCICT)* (pp. 1–4). Mumbai: Ieee. doi:10.1109/ICCICT.2012.6398180
- Spiess J, T’Joens Y, Dragnea R, Spencer P, Philippart L (2014). Using big data to improve customer experience and business performance. *Bell Labs Technical J.* 18(4):3-17
- Strydom J (2015). David Against Goliath: Predicting The Survival Of Formal Small Businesses In Soweto. *Inter. Busin. Econ. Res. J. (IBER)*, 14(3):463-476.
- The Retail Industry on the Rise in South Africa. (2012) (pp. 1–39). Retrieved from <http://www.treasury.gpg.gov.za/Document/Documents/QB1>
- The Retail Industry
- Ularu E, Puican F, Apostu A, Velicanu M (2012). Perspectives on big data and big data Analytics. *Database Syst. J.* 3(4):3–15.
- Yadav R, Kumar T (2015). Usage of Big Data Analytics for Customer Relationship Management. *Inter. J. Adva. Res. Computer Sci.* 6(2).

## Appendix A: Semi-structured interview questions

The interview questions were structured to cover themes and topics identified in the literature.

### General

1. What is your professional background? (qualifications and work experience)
2. Describe your current role and position within the organisation.
3. What is your past experience on analytics and big data analytics projects?
4. How relevant is big data analytics to your job function?

### Big data

5. How would you define the term big data and what data attributes does it have?
6. Does your organisation currently deal with 'big data'?
7. How does your organisation currently measure the volume of data which they are storing (Terabytes, number of records/transactions)?
8. What is the approximate total data volume that the organisation manages for analytics, both today and in three years?
9. What different types of data is your organisation currently storing as big data for analytics and from where is the data being obtained? (structured, unstructured, semi-structured)
10. How and where is this data being managed for analytics?
11. How frequently is data generated (real time, batch) and how quickly is data delivered into a dataset for analysis?

### Big data analytics

12. How would you define the term big data analytics?
13. Is your organisation using/considering using big data analytics today and how is it used? Is it a mature user of big data analytics?
14. Describe the process of analysing big data.
15. What factors (internal and external) will motivate/motivated the use of big data analytics?
16. Where is big data analytics being used, who uses it in the organisation and for what?
17. Who owns or controls big data analytics in the organisation?

### Value of using big data analytics in retail

18. Is your organisation considering/currently using big data analytics to help with functions such as inventory management, customer sentiment analysis, and customer micro-segmentation and price optimisation?
19. What is your perception of the value of using big data analytics?
20. What are the main benefits which your organisation has achieved from using big data analytics? How else can big data analytics be applied to obtain benefits?

### Challenges in using big data analytics in retail

21. Does your organisation have to consider challenges such as privacy of information, scalability of analytical algorithms and lack of available analytical skills in using big data analytics? How are these challenges being managed?
22. What would you consider to be the main challenges encountered in using big data analytics? How are you handling these challenges?
23. What were the top barriers experienced in implementing big data analytics?
24. How were these barriers addressed?

### Big data analytics techniques and technologies

25. Is your organisation using techniques such as machine learning, predictive modelling and cluster analysis to help with analysing big data? Are technologies such as cloud computing and data ware houses being used for big data analytics?
26. What other big data analytic techniques and technologies are being used to analyse big data both today and in three years? (predictive modelling, machine learning) What are the motivations for using those techniques and technologies?
27. How are these techniques and technologies being used to analyse big data?
28. What other techniques and technologies can be used for analysing big data?

**Vendor products for big data analytics**

29. Is a big data platform being used to handle the analytics of big data?
30. What product is being used and what were the motivations for using that product?
31. Describe the underlying technology infrastructure that supports the big data analytics solution.
32. What big data vendor was used to obtain the product? Describe the implementation process and the challenges that were faced.
33. Do you have a big data analytics platform as well as other analytics platforms? Are they integrated?
34. When do you consider replacing your current analytics platforms and what would drive you to replace them?

**Concluding questions**

35. What is the future of big data analytics?
36. Do you have any final comments about using big data analytics within an organisation?
37. Do you have any recommendations for the interview structure?