IT PRACTICES FOR SME SUCCESS SERIES

BOOK 3: BIG DATA FOR SMEs: QUESTIONS OF OPPORTUNITIES, CHALLENGES, BENEFITS AND OPERATIONS

2ND EDITION
2013

EDITED BY LIAM MAWHINNEY AND RICHARD SELF
PREFACE

This book is the 2nd of a continuing series of publications, composed of 45 articles and written by A grade students at the University of Derby. These students have been taught the technologies and practices that are at the cutting edge of enterprise systems.

The focus of this book is Big Data for SMEs which is a growing field with much contradictory information being presented to organisations of all sizes and sectors. The aim of this book is to critically evaluate the key questions of the opportunities, challenges, benefits and operations that need to be addressed by SMEs as they consider the business potential of becoming involved in Big Data Analytics projects with a view to extracting business value.

This book has been written for all levels of management and members of staff or users working in small to medium enterprises where information systems have an impact, or are integral, to their business. The articles written have been done so with this in mind, and as such, any recommendations and statistics are centred around these types of businesses.

It aims to inform readers on many of the most applicable business practices and strategies that should be considered with regards to Big Data Analytics systems (IS), thus increasing awareness in the workforce, allowing optimisation of system practices and ultimately driving success in SME’s, delivering more of what they should, in the way that they should.
# Table of Contents

Exploring The Benefits of Using Big Data When Creating Mobile ................................................................. 7  
Mohammed Saeed S. H. Abdulla  

Big Data In Marketing ............................................................................................................................................. 12  
Farah Ahmed  

Applying Big Data For SMEs .................................................................................................................................. 16  
Mohammed Al Yousif  

Big Data And It’s Value To SME Restaurants .............................................................................................................. 20  
Awfa G S Al-Adawi  

Big Data Analytics ........................................................................................................................................................ 24  
Hassan Al-Emadi  

Big Data Analytics: What A Small To Medium Enterprise Should Know ................................................................. 28  
Gareth Allen  

Big Data Analytics For SMEs ........................................................................................................................................ 33  
Maha Alshahri  

Opportunists For Supermarkets .................................................................................................................................. 38  
Adel Alsooj  

Big Data For Insurance .................................................................................................................................................. 42  
Ali Al-Suwaidi  

Implementation and Data Latency Issues Regarding Big Data ......................................................................................... 47  
Matthew Bateman  

Parallelization Considerations For Big Data Analytics ................................................................................................. 52  
William Briggs  

Big Data On A Small Scale ............................................................................................................................................ 57  
Adam Charlton  

Big Data In The Music Industry ..................................................................................................................................... 61  
Alex Clark  

Big Data As An Opportunity for the Tourism Industry ............................................................................................... 66  
Marcin Tomasz Drozdz  

Big Data for Retail SMEs ................................................................................................................................................ 71  
Oliver Fox  

The Challenges of Big Data ........................................................................................................................................... 75  
Mohammed Haroon  

Overcoming the Talent Shortage in Big Data Analytics ................................................................................................. 80  
Robert Heeley  

How Green is Big Data Computing? ............................................................................................................................... 85  
J. Hill  

Big Data Analysis: Is Web Analytics Sufficient for SMEs? .......................................................................................... 90  
James Hunt
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right To Privacy</td>
<td>94</td>
</tr>
<tr>
<td>Edidiong Emmanuel Inokotong</td>
<td></td>
</tr>
<tr>
<td>Big Data Breaches</td>
<td>99</td>
</tr>
<tr>
<td>Samuel James</td>
<td></td>
</tr>
<tr>
<td>Can SMEs Use Big Data To Help Make A Successful Game?</td>
<td>103</td>
</tr>
<tr>
<td>Michael King</td>
<td></td>
</tr>
<tr>
<td>Big Data and Recruitment</td>
<td>110</td>
</tr>
<tr>
<td>Vasiliki Kozi</td>
<td></td>
</tr>
<tr>
<td>Big Data, Big Lie</td>
<td>114</td>
</tr>
<tr>
<td>Daniella Kypri</td>
<td></td>
</tr>
<tr>
<td>Big Data Opportunities</td>
<td>119</td>
</tr>
<tr>
<td>Evaldas Luksys</td>
<td></td>
</tr>
<tr>
<td>The Opportunities and Challenges of SMEs Using Big Data within the Travel and Hospitality Industry</td>
<td>125</td>
</tr>
<tr>
<td>Liam Mawhinney</td>
<td></td>
</tr>
<tr>
<td>Security Issues with Big Data within SMEs</td>
<td>130</td>
</tr>
<tr>
<td>Amrith Nagra</td>
<td></td>
</tr>
<tr>
<td>Multimedia Big Data</td>
<td>135</td>
</tr>
<tr>
<td>Umair Farooq Naru</td>
<td></td>
</tr>
<tr>
<td>Big Data Security Challenges in SMEs</td>
<td>140</td>
</tr>
<tr>
<td>Feyisayo Obisesan</td>
<td></td>
</tr>
<tr>
<td>Big Data in Retail Supply Chains</td>
<td>145</td>
</tr>
<tr>
<td>Marialena Panagiotidou</td>
<td></td>
</tr>
<tr>
<td>Life Science: Genomic Sequencing in the Era of Big Data</td>
<td>150</td>
</tr>
<tr>
<td>Gwyn Dafydd Owen Perkins</td>
<td></td>
</tr>
<tr>
<td>Big Data within SMEs</td>
<td>156</td>
</tr>
<tr>
<td>Simon Ranson</td>
<td></td>
</tr>
<tr>
<td>Moral and Ethical Ramification of Collecting Big Data from Consumers</td>
<td>160</td>
</tr>
<tr>
<td>Joe Rawlings</td>
<td></td>
</tr>
<tr>
<td>Big Data: SMEs Biggest Challenge</td>
<td>164</td>
</tr>
<tr>
<td>Joao Vasco Zurzica Reis</td>
<td></td>
</tr>
<tr>
<td>How Web Developers Can Use Trends Data to Provide Big Data Services to E-Commerce Clients</td>
<td>170</td>
</tr>
<tr>
<td>Tim Ride</td>
<td></td>
</tr>
<tr>
<td>Big Data fro Search Engine Marketing</td>
<td>174</td>
</tr>
<tr>
<td>Stephen Ridgway</td>
<td></td>
</tr>
<tr>
<td>Big Data for SMEs</td>
<td>178</td>
</tr>
<tr>
<td>Kareem Samarah</td>
<td></td>
</tr>
<tr>
<td>Big Data for SMEs: Questions of Opportunities, Challenges, Benefits and Operations</td>
<td>182</td>
</tr>
<tr>
<td>Tomas Satala</td>
<td></td>
</tr>
<tr>
<td>The Power of Big Data in the Retail Industry</td>
<td>187</td>
</tr>
<tr>
<td>Umar Shakil</td>
<td></td>
</tr>
</tbody>
</table>
Abstract—This paper provides a summary of how big data can be useful for small and medium enterprises in the 21st century. Organisations can draw meaningful information from analysis of how large amounts of data are created on a daily basis, which can help them make informed decisions. Case studies in which SMEs have successfully utilised the potential of big data are cited in the paper. The paper shows the potential of big data for mobile application development enterprises, which fall under the SME bracket, and how they can utilise the benefits of big data, by presenting applications-based analytical results. The conclusion presents a syllogism whereby a mobile app development company has succeeded in using big data for generating revenue streams, and therefore provides an example of how big data could be utilised by other firms.

Index Terms—Smartphone, application, big data, social media, market.

I. INTRODUCTION

The rapid growth of the Internet, smartphones, social media, wireless technologies, etc. in the digital world has led to an explosion of data. It is predicted that the next technological revolution will be based upon the science of ‘big data’, through which large amounts of data can be processed, captured, stored, shared and analysed. According to estimates by analysts, these large quantities of data will increase by 40 times in the upcoming decade (Intuit 2020 Report 2012).

Previously, the ability to gather and capitalise upon large amounts of information was limited to large enterprises, since they possessed a pool of statisticians who could obtain meaningful information from the data. However, due to the democracy provided by big data, small and medium enterprises can utilise the benefits of data analysis, which will help them to obtain meaningful insights into competition, markets, bottom lines and top-line results, and enhance their decision-making abilities (Brown and Duguid 2002, p. 29, 39).

The ease of access to large amounts of data has made data a vital resource, along with labour and capital, in any industry. As predicted, the key driver for growth in the 21st century for any organisation, big or small, is data that encompasses many societal aspects, such as health care, business, entertainment, government, finance, etc. (Erbes et al. 2012, pp. 66–72).

In collaboration with Emergent Research, Intuit, an American software company, analysed and predicted data trends for the next seven years. The report estimates that 30 zettabytes of data will be stored in the year 2020 (Intuit 2020 Report 2012). IBM discovered that around 10 terabytes of data are generated daily by Facebook, and 7 terabytes of data are generated daily by Twitter (Intuit 2020 Report, 2012). Similarly, McKinsey predicts that data will grow 40% annually, while CISCO concludes that the number of mobile-connected computing devices will exceed 10 billion by 2016 (Manyika et al. 2011, p. 4), an estimate that has grown from under 7 billion in 2012 (Intuit 2020 Report 2012).

Gartner, an IT and research firm, has analysed and graphically presented the results of over 2000 technologies, grouped into 98 categories, that outline and depict the life cycle of technologies which have a high potential (Ross 2011). Technology is expected to stabilise during the later stages of its life cycle, since by that time society and industry will accept it as indispensable, and will not create undue hype. In this research, high expectations for the future are placed upon big data (Gartner 2013).

For small and medium enterprises, this data revolution promises a number of benefits that will enable them to make more informed decisions, improve communities and open new avenues for business opportunities. As considered from the perspective of a mobile application development company, the potential for the huge amounts of data that will be created by smartphone technologies can be utilised through big data, and can create more useful mobile applications for the market (Erbes et al. 2012, pp. 66–72).

II. DEFINITION

Big data is a term that covers various aspects of data, ranging from technological bases to economic models. NESSI defines big data as ‘A term encompassing the use of techniques to capture, process, analyse and visualise potentially the largest datasets in a reasonable timeframe not accessible to standard IT technologies’ (Big Data: A New World of Opportunities 2012, p. 6, 18).

Through this definition, NESSI introduces big data as a combination of tools and techniques that works beyond the limits of normal data-interpreting IT tools. Although big data tools use IT tools to run analyses, the results and inferences obtained from big data will be more complex than the simpler IT software (Big Data: A New World of Opportunities 2012, p. 6, 18).
Big data is also defined by PWC-Australia in the European Commission’s report: ‘Big data is a high volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making’ (European Commission 2012).

This definition explains that big data is a large amount of data that cannot be stored, captured and analysed by traditional means for accurate and timely decision-making purposes. Because of the sheer volume of data, and because it is increasing exponentially, the data can be sorted only by dividing it into categories. According to the European Commission (2012), all large amounts of data that must be processed by big data can be categorised as:

- Internal data
- Unstructured external data
- Structured external data

All the above types of data can be utilised by small and medium businesses for generating profits.

Fig. 1. Institutes Start Using Their Internal Data, And Then Learn To Obtain And Structure External Data (PWC, 2012, P. 4, 6)

A. Internal Data:

Internal data can be any data produced by a business through its own system and tools during its daily operations. This data can consist of both qualitative and quantitative information that can be used by the business to gain insights into its internal processes. This data can be used by small and medium enterprises to improve their existing processes, streamline their functions and reduce their costs.

B. Unstructured External Data:

This type of data is generated from sources and means outside of a business. Something as simple as a Google search can be considered to be unstructured data, since the algorithm will take into account over 200 factors while populating results.

C. Structured External Data:

This type of data is often provided to a business by external sources or vendors. For example, marketers use sales data from Walmart to better understand customer preferences (Manyika et al. 2011, p. 4).

III. Analysis

A. How Are Big Data Analytics Beneficial?

Big data falls under the umbrella of ‘business intelligence’, in which it plays a major part. Big data analytics can provide critical business intelligence on customer behaviour and trends (Centre for Economics and Business Research Ltd. 2012, p. 4). Predictions of the buying patterns of customers can be made with increased accuracy using big data. As a result of such benefits, the use of analytics for decision-making is becoming inevitable. Companies have used the potential of big data to improve their performance and yield greater profit margins (Gualtieri 2013).

B. Market Analysis of Big Data

Big data analytics could be applied to various sectors, such as retail, health care, media, utilities and telecom. Various techniques for gathering, processing, analysing and visualising data have been used in each sector (Floridi 2012, pp. 1-3). Market analysis shows that the visualisation of data is as important as the analysis done on it. Currently, the major players in the market are SAS, IBM, Vertica, Opera Solutions and Mu Sigma, closely followed by Aster Data, Splunk, Greenplum, etc. (Probst 2013). Despite these market players, SMEs have emerged with proprietary solutions that use big data analytics in order to enhance their business and obtain positive results (Probst 2013).

C. Analytics and Marketing

The trend of using analytics for marketing purposes is in its initial stage, and companies have not fully utilised the potential of big data for marketing and market analysis. However, big data marketing has many benefits. The decision-making capability of firms is enhanced by the use of big data; it increases accuracy, time management and customer segmentation, and allows for real-time marketing. The continuous increase of data is predicted to continue and grow by 60% for marketing analytics in the next three years (European Commission 2012). According to research by IconicMind.com, less than 10% of marketers are using data-driven marketing techniques (Mudd 2013). However, managers are making a conscious decision to implement big data that will cover 71% of the marketers in the business using big data for decision-making. In such a market, where scenarios are adopted for big data, SMEs cannot afford to miss out on the wave of data-driven decision-making and data analytics for marketing purposes (Mudd 2013).

D. Productivity and Sales Increase

The use of big data for decision-making can help a firm to increase its productivity by an average of 10%, according to research by the University of Texas and Sybase. The same research states that the use of big data has led to increased productivity in the retail industry by 49%, resulting in a $1.2 billion increase in sales revenue (Sybase 2013). Similarly, consulting and telecommunications industries have seen an increase in productivity of 39% and 17%, respectively, and an
increase in sales revenues of $5 billion and $9.6 billion, respectively (Sybase 2013).

E. Big Data for Mobile Data

The volume of mobile data traffic is estimated to increase to 10.8 exabytes per month by 2016, according to research by Forrester (Schadler and McCarthy 2012). The majority of this exchange of data is possible due to applications created across iOS, Android and Windows platforms. Big data technology is required to comprehend such large amounts of data being transmitted over mobile networks. Application developers can use this data to gain insight into the types of data being transmitted by big data (Laurila et al. 2012, p. 1).

IV. EFFECTS OF 3VS OF BIG DATA

A. Velocity

Big data velocity deals with the pace at which data flows from various sources, such as IT systems, business processes, networks, etc., since most of the potential benefits of big data for businesses are dependent upon real-time data analysis. This factor plays a crucial role when recognising success due to big data (Normandeau 2013).

B. Veracity

Big data veracity refers to noise and abnormality in the data, and all the data elements not required for analysis. The principle ‘Garbage In – Garbage Out’ applies to big data because it deals with large volumes of data. If such volumes are not clean, and if they have biases or abnormalities, the accuracy of the results is affected (Normandeau 2013).

C. Validity

Big data validity refers to the correctness and accuracy of the data considered for analysis. Veracity and validity go hand-in-hand in big data analytics, since any deficiency related to these two factors will yield less valued and incorrect results (Normandeau 2013).

V. CASE STUDY

The following cases show how big data analytics have been used by small and medium enterprises to gain competitive advantages in their respective industries.

A. Digital Route Data Processing for Telecommunications Industry

Breakdowns in the systems of telecom companies disrupt the seamless, real-time flow of data used for billing telephone users. To ensure this seamless flow, mediation software solutions are used to avoid data loss, which can directly affect revenues (Thomas and Cook 2005).

Digital Route, a medium-sized enterprise, devised software called ‘MediationZone’, which uses big data analytics to measure data traffic and provides real-time analysis to the telecom operators so they can understand which servers are overloaded due to high data transmission. With this information, timely decisions can be made, and systems can be protected from disruption (Probst et al. 2013, pp. 6–9).

B. Neodata’s Advertising Insight

Increasingly, publishers and media companies are relying upon online publications to reach audiences. Since the majority of online publications are free, advertisement becomes the only source of income. Advertisers are inclined to place their ads next to the most frequently read articles in order to increase sales or hits on their website (Probst et al. 2013, pp. 6–9).

Neodata, a small-sized enterprise, offers a solution which uses semantic analysis, text mining and real-time web metrics to allow publications to analyse the most-read article on their website, which allows them to align relevant advertisements next to the article and earn revenue (Probst et al. 2013, pp. 6–9).

C. Treemetrics

Treemetrics, an Irish start-up, is another small firm that has established itself as the Google of forestry, with data analytics for over 11 million trees. They use big data analysis to help optimise harvesting and reduce waste (Probst et al. 2013, pp. 6–9).

D. oDesk

An online talent marketplace, oDesk.com, uses big data predictive analysis to improve the matching of employers with freelancers. It draws from a database of fully completed assignments, then simplifies the task of evaluating and selecting part-time, which increases the chances for project success (Intuit 2020 Report 2012).

VI. CONCLUSION & RECOMMENDATION

Mobile application development companies, which typically fall into the SME segment, can use the same strategy as other SMEs in their industry to utilise the potential of big data. The biggest advantage a mobile application company can gain from big data is insight into which content is accessed from mobiles for which industry. Based on such insights, it is possible for a mobile app development company to design its product for the market (Viktor and Kenneth 2013).

IMS, a U.S.-based big data and analytics company, recently launched its health app, which helps over 40,000 users monitor their daily health (IMS Health 2013). IMS serves as an example of a mobile app company developing products and making product decisions based on data processed by big data in order to maximise its yield (Intuit 2020 Report 2012, p. 1, 13).

Since big data provides real-time analysis, the industry’s requirement for real-time reports is a high priority. As a result, companies and decision-makers need real-time analysis to be available on their mobile devices. These kinds of demands provide mobile application companies with a gateway to develop applications that can deliver big data on their clients’ mobile devices (Raskin 2000).

Estimates predict that 44 billion mobile applications will be downloaded by 2016. This provides insight into the growing opportunities in the field of mobile app development (Allied Business Intelligence Inc. 2011). The results delivered by big data can help mobile application development companies
decide which types of applications are most highly used, enabling the developers to create similar but improved applications that will eventually increase their revenue streams. In addition, big data is predicted to increase the employment ratio in the mobile app industry (Lavalle et al. 2010). Hence, the number of mobile application developers will increase due to employment opportunities, resulting in a constant flow of human capital for mobile app companies (Davenport et al. 2012).

VII. REFERENCES


Big Data In Marketing

Farah Ahmed
School of Computing and Mathematics
University of Derby
Derby, Derbyshire, United Kingdom
F.ahmed1@unimail.derby.ac.uk

Abstract—This report summarizes the effects of Big Data and how marketing in Big Data also has its own affects. The sheer volume of information that is available out there is at such high degree, that it would not be a surprise if marketers want to ignore it. However, for those marketing companies that are brave enough to confront big data there are continuous rewards. Seven different definitions of Big Data will be introduced in this paper. It will also critically evaluate big data in marketing and its advantages and disadvantages. There are five V’s that have been used for the purpose of this study and their definitions and consequences will also be delivered. Following this, at least one question regarding big data in each, Opportunities, Challenges and Benefits sections will be given and then finally this report will come to an end with the Conclusions and recommendations.

Index Terms—Big Data, Marketing, Technology, Value, Validity.

I. INTRODUCTION

There are many definitions of ‘Big Data’. Elliott (2013) explains seven definitions of big data in his report in Business Analytics; The Original Big Data, Big Data as Technology, Big Data as Data Distinctions, Big Data as Signals, Big Data as Opportunity, Big Data as Metaphor and Big Data as New Term for Old Stuff.

‘The Original Big Data’, is where Elliott (2013) describes this popular definition as the “three Vs”. Volume, Velocity and Variety which was first created by Doug Laney of Gartner over twelve years ago. (Elliott, 2013).

‘Big Data as Technology’ is an old term of big data which was driven by new open source technologies of storing and manipulating data. Along with the usage of the new tools, a new term was created to separate them from previous technologies, so the unique and of course obvious term created is ‘Big Data’. (Elliott, 2013).

‘Big Data as Data Distinctions’ helps organisations understand how Big Data is different in this decade than Big Data in the previous decade. Transactions, Interactions and Observations. Transactions is most of what has been collected stored and analysed in the past. The interactions is the data that arises from people clicking on webpages, and observations are the data that is collected automatically. (Elliott, 2013).

‘Big Data as Signals’ is a slight “business approach that divides the world by intent and timing rather than the type of data” (Elliott, 2013). The data from the past is about transactions, and by the time these transactions are recorded there is not much an organisation can do about them, however, the new Big Data organisations can use new signal data to expect what will happen and step in to improve the situation.

‘Big Data as Opportunity’ is a broad definition that states big data as “analysing data that was previously ignored because of technology limitations.” (Elliott, 2013).

‘Big Data as Metaphor’ is defined in the book ‘The Human Face of Big Data’ by Rick Smolan as “the process of helping the planet grow a nervous system, one in which we are just another, human, type of sensor.” (Elliott, 2013).

‘Big Data as New Term for Old Stuff’ is basically saying that the new Big Data is a lazy and cynical term for projects that were possibly using former technology.

Overall, the huge debate over the definitions of big data still exist today, yet they all agree on one thing; that big data is a massive deal, especially in this day and age, and will lead to fantastic new opportunities in the future.

II. BIG DATA IN MARKETING

To enhance marketing responsibilities big data has been widely organised, however the small amount of success stories show that there may be a significant number of failed applications that have been put into practice. (Canopy, 2013). There are reasons being such strong enthusiasm behind Big Data by Chief Marketing Officers (CMOs), one of a few is that they have responsibilities that are alongside Big Data capabilities. The figure below shows CMOs core responsibilities and Big Data use case examples, this is the reason why CMOs will carry on using Big Data to deliver actionable perceptions.

Fig. 1. CMO Responsibilities And Big Data Use Case Examples (Canopy, 2013).
A. Advantages

Many Chief Marketing Officers (CMOs) place Big Data in their top priority list and nearly three quarters of marketers plan to have control over Big Data analytics solutions within two years. (Canopy, 2013). Big data requires new roles and skills, but because Big Data is hugely popular today, soon there would be a ‘do-it-yourself analytics’ which would make it much more widespread and will also be very valuable to the public. (ESDS, 2013).

B. Disadvantages

According to Kotorov (2013), the real disadvantage of ‘big’ is the high risk of failure, and another disadvantage is failing to take the risk and being downgraded by those who did. “Damned if you do, damned if you don’t!” (Kotorov, 2013). A third disadvantage of ‘big’ is the ‘wait-and-see’ mind set. This is where an organisation waits to see their competitors implement Big Data, analyse and learn from their mistakes and then implement their own project when time is convenient.

C. Concluding Big Data in Marketing

Big data in marketing can generate substantial value for Chief Marketing Officers (CMOs) and their organisations when planned and employed correctly. (Canopy, 2013). This is where first steps come in, for example, knowing the risks of Big Data before implementing it. As the focus of Big Data is warming every day, the expansion in the volume and variety of data increases and the vast speed of generating new data is accelerating rapidly. Overall, the rise of Big Data is highly positive for society in all aspects. (ESDS, 2013).

III. The Five V’s

A. Volume

The amount of data of the volume of data that is to be analysed. This is a vital thought in determining the need for Big Data in an organisation as standard database systems may not be sufficient enough to handle large volumes of data. (Canopy, 2013).

B. Velocity

This is the speed at which data needs to be created, processed and stored very quickly as the data may be time-sensitive. (Big Data – Startups, 2013). An organisation will also need to determine the need for Big Data when analysing the stored data as previous hardware may fail to meet these requirements. Using Big Data in marketing looks to use social media data streams to improve the customer relationship management in an organisation, this can often benefit pricing and new product development. However, an organisation using social media may need to handle a vast amount of inflows of data each day. (Canopy, 2013).

C. Variety

Variety covers the different forms of data i.e. structured and unstructured which contain images, emails, spreadsheets, social media conversations and streaming media. So far, there is no ‘one-size-fits-all’ approach (Rob Livingstone Advisory, 2013). So as unstructured data increases, the difficulty for marketing tools to deal with this increase as they only work with structured data, so marketing efforts need to influence different data sources to generate consumer profiles. (Canopy, 2013).

D. Validity

Validity of data is interpreted having a complete judgement or fact, this is the result of the logical implications from matching data. (Rob Livingstone Advisory, 2013).

E. Value

The value of data shows its importance, data in itself has no value at all, however, having data that is worthy and useful is highly relevant to an organisation. (Rob Livingstone Advisory, 2013). Here Big Data means big business and every industry will earn the benefits of Big Data. The real value of data is in the analysis that is done and how the data is then turned into useful information, which would then turn into knowledge. Organisations can turn this valuable knowledge and turn their organisation around, this would be highly beneficial for decision-making. (Big Data – Startups, 2013).

There are reasons why we do Big Data Analytics, and one is that the large volume of data that comes in from the public makes the information so valuable, that some people are already seeing a difference, this would then result in other companies wanting to follow. Examples of this are given in the Benefits section.

IV. Questions

A few questions should be asked regarding Big Data in each of the sections below.

A. Opportunities

The opportunities that Big Data makes possible are in sheer volume, this is due to the fact that organisations are now aware of the fact that using data that’s out there is a potential asset to them. (SAS, 2014).

According to Taylor (2014), Eric Schmidt, Google’s former chief executive, said that in 2010 approximately five exabytes of data, which is the same as 250,000 years of DVD quality video, was created in the world every two days. There are estimates that next year the same amount of data will be created every ten minutes. (Taylor, 2014).

We are in a highly advanced statistical processing ‘machine-learning, where with more data in this cynical model, the smarter the systems. Some companies are adding data collected from customers using their smartphone and tablet applications. There are other companies that go through vast volumes of social media traffic that is generated every day to look for consumer trends. Taylor (2014) states that research vice-presidents at Gartner, Mark Beyer and Doug Laney say, “Retailers are attempting to create ‘graphs’ of social networks…to create social buying patterns”. More and more companies are buying other types of data, for example, weather data, traffic information or website statistics. This is so they can build a more comprehensive picture of their customers.

This shows that the real business opportunity is in the ability to put more data together and let the data sources support each other. This way, big data makes organisations
stronger, smarter and more productive, this is by allowing people to join different data types that were unavailable before, and enabling them to discover opportunities that have never been seen before.

The ‘big’ question to ask is; Will the opportunities that Big Data creates ever come to an end?

B. Challenges

There may be many advantages in implementing Big Data, however according to Canopy (2013), it is likely that CMOs would face three common challenges.

1) Lack of Alignment on the Objectives

In order to keep up with the current trends, many companies want to implement ‘something’, ‘anything’, as long as it is with Big Data, doing this may focus on the technical aspects of the solutions but can fail to meet the business needs.

2) Confusion About When Big Data Is Truly Required

To solve many data analysis problems, many turn to big data instead of checking if existing capabilities can actually be achieved.

3) Difficulties of Technical Implementation

Data takes a significant amount of time to prepare and standardise before analysis, and very few CMOs are data scientists or have the high expertise or knowledge to do this so a high level of effort would need to be invested into training to integrate new tools into day-to-day workflows. However, because the industry is clearly moving towards making Big Data compatible with BI interfaces, there would be a requirement to hire highly experienced data scientists, a Big Data implementation should be complete and appropriate training and tools would be required. This requirement is not only for the department that is implementing Big Data but also for everyone that is influenced by the requirement and the benefits. (Canopy, 2013).

Taylor (2014), states that Hasso Plattner, the co-founder of SAP says that “data has proliferated at unbelievable speed,” he then goes on and points out that there are two types of data that are being captured, ‘structured data’ which is easily analysed, and ‘unstructured data’ which are mainly social media updates and videos etc. these are much harder for computer to decode.

Another challenge for big data is when organisations try and use social media to match up social networking data with very well structured and validated data about customers. Organisation don’t often collect social media comments, updates or tweets with much validation. Even sorting and matching data by names or D.O.B is a huge challenge and can incur in a high percentage of mismatch. This is because so many people share the same names and D.O.B. Mismatch in an organisation can cause serious problems, especially in financial companies. For example, if a loan company refused someone a loan then found out afterwards that their name was matched with somebody else’s. This causes customer relation issues and the innocent individual receives bad credit for no reason.

According to SAS (2014), there are three additional challenges for marketing; knowing what data to gather, knowing which analytical tools to use and knowing how to go from data to insight to impact. Three questions derive from this:

- Are you aware of what type of data to gather?
- Do you know what analytical tool you will be using?
- How are you going to use the insight to make a positive impact on your marketing programs?

C. Benefits

Out of all the benefits that Big Data offers an organisation, according to SAS, the largest would come from “improved business efficiency” (SAS, 2014). While Big Data will drive innovations with an estimation of £24 billion in value. New opportunities for SME’s are predicted to be worth £42 billion. The industries that will benefit the most would be retail banking, manufacturing and retail. The government can also save £2 billion on fraud detection and £4 billion through improved performance management. (SAS, 2014).

V. CONCLUSIONS AND RECOMMENDATIONS

In conclusion, Big Data in marketing shows to be a huge and continuous success. The only disadvantage that has derived from the research conducted is the lack of data management expertise or the lack of management skills in general to pursue anything with big data. If this is the case in some organisations, then it would be a real shame as there are too many companies out there already that are missing out on great opportunities. Other organisations are missing out because they still attempt to capture, manage and process data in traditional database systems. If an organization does not upgrade their systems, then how can they expect their business to upgrade?

Retail is one of the sectors that has said to have the greatest potential for big data. This is due to their customer transactions, both online and offline. This can all be brought together, including conversations and intentions, so that brands can understand how to reach their customers better.

Organisations that use social networking data to match with their validated internal data do so at high level of risk. There are other ways in improving values and customer satisfaction without trying to match very well structured data with unstructured data.

When an organization are looking at social media results, comments, updates, feeds and tweets etc. and find some negative feedback, what they ask themselves is; “how much can we trust the data and how representative is it of our customer base?”. What companies at this stage say to themselves is, “well, it is not constructive to consider ALL negative twitter tweets, however, they do display warnings of problems where some people are getting very irritated, angry or upset by some of the things we are doing, so therefore, we should fix these problems mentioned by the small tweets or postings,” this would then make the overall customer experience better, and that of course leads to improved customer satisfaction, which then of course leads to improving value.

Sometimes size does not matter, as long as the data, even small data, is the correct data that helps solve a problem or
address a question, that is what matters. (Marketing Pilgrim, 2013).

VI. REFERENCES


Applying Big Data For SMEs

Mohammed Al Yousif
School of Computing & Mathematics
University of Derby
Derby

Abstract—To apply big data on SMEs, there are some elements that need to be considered. Trends manipulation and being focused on other things – instead on ‘big’ aspect of data itself – are very useful. Data that already exist can create valuable opportunities. Social media increases the chances for SMEs to grow quickly at low cost. Velocity and Variety is important for SMEs. Big data requires strategies to make it beneficial instead of a liability.

Index Terms—Big Data, Social Media SMEs, B2B SMEs, Vs, Velocity, Variety.

I. INTRODUCTION

“It is a capital mistake to theorise before one has data.” – Sherlock Holmes

Current enterprises use data with external data to expose new visions through use of special service or software. Unlike big vendors, SMEs can gain access to many Web options, such as cloud-based options to help analyse data using analytic clouds, such as Google Analytics. With these analytical advantages, SMEs will gain superb performance in creative ways (Simons, 2013). With social media, there is no need for a big budget or big name brand. Online marketing creates engagement to the customer base (Shuen, 2014). According to Enterprise Research vice president Alexandre Mesguich, many vendors do not really understand what ‘big data’ is and do not know its abilities or how it could support technology interaction through current platforms (Macinnes, 2013). Even though big data are there, how could it be useful for small and medium enterprises SMEs?

II. SMES & BIG DATA

Group and marketing director at Six Degrees Group, Campbell Williams, stated that SMEs do not have a market for big data because they do not have big data. If SMEs work in such a pattern, they would need what is required to develop and grow (Macinnes, 2013). Roambi’s Dave Becerra, vice president of strategy and business development, has a different point of view: He thinks most SMEs have started to realise the value of big data and found it bigger than just a direction (Macinnes, 2013). They now understand that big data can help gather useful information, such as competitive information and identifying target markets (The Big Data Insight Group, 2012). Even though SMEs have small resources and smaller resources, they have many other qualities to compensate, such as flexibility with IT substructure, the ability to change performance swiftly, and fewer system issues or dissimilar databases (Simons, 2013).

III. THE V’S AND SMES

To achieve successful big data results for SMEs, the V’s need to be considered as a main dimension of success. There originally used to be only 3 V’s; today, there are at least 11 V’s that can be applicable to SMEs. The original 3 V’s are volume, variety, and velocity. Volume = data amount. Variety = the number of data types. Velocity = how fast data are processed (Rouse, 2013). The 11 V’s are: volume, velocity, variety, variability, value, veracity, validity, volatility, verbosity, vulnerability, and verification. There can be more or fewer than those V’s depending on need and an enterprise’s perspective.

Data trends seem unstoppable (Industry Perspectives, 2012). When big data appear, the 3 V’s come with them. As a result of data increases, the speed of data processing needs to be faster. General transactions create a veritable continent of valuable transactions, such as credit cards and phone calls (Industry Perspectives, 2012; Information Management Online, n.d.). The safety of information and keeping it moving, even when it is distributed, are other important aspects of big data (Industry Perspectives, 2012).

A. Variety is Important

Variety implies an assortment of sources and types of data (Trader, 2014). It is not necessary for data variety to be in structured form; it could also occur in the form of social media text or an image (Media, 2012). When it comes to big data, variety is more difficult to handle than velocity because variety requires more focus and experience with the subject itself. Variety is concerned with putting data into processed shape; therefore, variety must reflect a sound understanding of the sources when data are processed (Barash, 2014). While technology can effectively handle both volume and velocity, variety is still anti-software (Trader, 2014). It is very difficult to work out variety using programs (Kumar, 2014). Rather, variety requires big data experts in certain applications and domains. This renders the outcomes of processing a variety of big data conditional upon the skills and expertise of the people who work with these data (Kumar, 2014). While variety is very important for SMEs, it requires specific skills to be processed and managed. SMEs need to be more careful when dealing with variety than they do when dealing with volume and velocity.

B. Velocity is Important

This particular V – velocity – is responsible for the speed of data movement from sources such as networks, social media, mobiles, and business processes. The high velocity here creates
more value, and it is important to extract the supreme value out of that data and then process it within a very short period of time (Oracle, n.d.). This data flow can create opportunities for business and researchers to create valued decisions that lead to advanced competitive improvement (Normandeau, 2013).

Data volume has been exploding: In 1986, the percentage of digital data did not surpass 1% of all data collected. A huge turn took place in the years that followed 2000. This period witnessed a data revolution and a dramatic rise, from 25% up to 94% of digital data in 2007. The factors that played the most significant roles in this data revolution were (a) the development of technology, (b) social media, and (c) CRM and ERP systems with predominance, gaining, and fusion (Chardonnens and Cudre-Mauroux et al., 2013, p. 784–787). Data speeds are increasing because of better computer systems; therefore, consumer numbers are growing because of that data. The growth is obvious from operational and analytical aspects (ScaleDB, 2014). Humans have been generating various kinds of data for centuries, but these newer processes have grown significantly: 90% of data globally was generated within last couple of years. This gives a clear the image of big data in the world. Nevertheless, both data created by actions of humans and devices were not traced or collated until recently. This is because technology was unable to process this voluminous amount of data with velocity (ScaleDB, 2014). Now that we have the tools to manipulate and analyse this load of data, people have found reasons to store these massive data quantities (ScaleDB, 2014).

C. Important High Velocity Data Sources for SMEs

- Log Files – websites, technology log events, devices, and database (ScaleDB, 2014).
- IT Devices – routers, networks, printers, or any devices that create useful data (ScaleDB, 2014).
- User Devices – smartphone activity is valuable and logged (ScaleDB, 2014).
- Social Media – massive amounts of fast, valuable data come from social media platforms, such as Facebook, Instagram, and Twitter (ScaleDB, 2014).
- Online Gaming – online games are valuable sources to find users’ interaction data, not only with the game itself but with other users (ScaleDB, 2014).
- SaaS Applications – the beginning of SaaS had some limitations of functionality. However, as they grew, the users’ association and connections grew, resulting in an enormous flow of data (ScaleDB, 2014).

D. How Does Velocity Increase Marketing For SMEs?

There are three important aspects to be considered in order in making velocity valuable for SMEs:

1) Static Data

Static data remains the same for long periods; for example, look-up data. Normally, these data do not require high velocity, so there is no need to consider a big data driver (Roberts, 2014).

2) Dynamic Data

Dynamic data are current data of existing customers that are updated and added, such as online orders (Roberts, 2014). Under this, there are two subgroups:

- Change: Regularly updated data, such as details and information of a given customer (Roberts, 2014).
- Growth: Regularly growing data, such as outbound communication details that would be added each time interaction happens with a customer (Roberts, 2014).

3) Continuous Data

Continuous data are machine data and social data supplied in a remarkable; therefore, the range of data is growing constantly. This growth results in the need to analyse, capture, and gain insight with faster drives and bigger storage (Roberts, 2014).

IV. MARKETING BENEFITS OF THE DATA FLOW

A. Automation

Constant data flow creates opportunities to store, collect, and analyse data. Automation helps control data velocity, creating opportunities, including:

- Automated Response Management: This is controlled with data available at the moment creating opportunities come across the channels, responding to current activities (Roberts, 2014).
- Automated Personalisation: With velocity, the chance to know your customers is very strong, especially regarding the volume. This creates valuable immediate personalisation through mobile, local devices, and desktops (Roberts, 2014).
- Automated Actionable Evaluation: These are techniques to check KPIs and metrics against preceding and present levels of marketing. The marketing process will show the present act and ensure that the budget is successfully allotted to respond (Roberts, 2014).

B. Improvement of Marketing Performance

With analytical comparison, the data used with other sources create a high range of review of market activities (Roberts, 2014).

C. Targeting

With available data and data provided, the decision to target – or, conversely, not target – comes according to the knowledge available most recently. The ability to efficiently use and store data as it is generated will reduce the impact of obtainability of present activity and behaviour (Roberts, 2014).

V. SOCIAL MEDIA PLATFORMS FOR SMES

With super brands and agencies in the same industry, it is difficult to grab customers’ attention through social media networks such as Facebook, Instagram, Twitter and YouTube (Hargrave, 2013). As soon as social media impact is gained, small- to medium-sized enterprises will have direct access to their customers’ databases, which create the ability to target what customers like and need based on their social media pages. In addition, targeting future possible customers, gaining
access to competitors’ information and profiles, trailing customers’ history – even before they have become customers – are all possible with a simple mouse click (Hargrave, 2013). Social media platforms have the advantage of setting small daily budgets to market via specific followers and certain posts. It is possible to select the sort of individuals one wants by location and kind of interest, and payments will be charged only on followers who click on the advertisement (Hargrave, 2013). Do not be intimidated by megabrands’ super-followers lists: It is more important to aim for people who are interested in what you provide. Five hundred engaged users are much valuable than 5000 random followers (Hargrave, 2013).

A. How can SMEs succeed using social media platforms?

- Create a relationship with the customer - Most customers already involved in social media platforms. This creates a wonderful chance to engage with the specific market and announce new offers and services (Shuen, 2014).

- Improve and adjust - As soon as you get engaged with customers, use a Feedback function to upgrade and develop services and products (Shuen, 2014).

- Gain new customers - The shadow of social media will increase the customer database, which will result in increased revenue (Shuen, 2014).

- Growing Web traffic - Use your service to offer a goal for those who are interested in your type of service (Shuen, 2014).

- Be competitive and keep up! - Your competitors will also be engaged through social media with customers; this means you need to work harder and do better to ensure you do not get left behind (Shuen, 2014).

VI. B2B DATA APPLIED ON SMES

Analysing and collecting data is aimed at extracting more income from consumers (Stock, 2014). There are two processes to complete big data collection and analysis. Many analytical procedures can be applied to data that already exists in the case of limited data collection (Stock, 2014). Get competitive advantage and new opportunities by working B2B data (Stock, 2014). What would SMEs gain from big data business data analysis?

A. Other corporations linked and analysed by B2B data consents:

- Get an overview image of patterns and leanings per consumer, per sector, and per industry.

- Trends manipulation will get accurate personalisation through new marketing and sales.

- Identify new customers’ needs that have not been realised before.

- Create new services and products and then target customer needs.

B. Instead of wasting time on the ’big’ aspect of big data, SMEs need to focus on:

- Using existing information and applying big data, business styles, and data analysis.

- After each customer interaction, B2B data lists need to be updated.

- Advance sales targeting and marketing efforts through B2B.

VII. CONSEQUENCES OF BIG DATA FOR SMEs

With big data, many opportunities can be created through privacy invasion, superb financial coverage, and saving money and time in useless opportunities (Jordan, 2014). The tools of big data are not ready yet because they evolve frequently, and they are not included in most universities’ curriculum. These tools have minimal vendor support. More user flexibility is needed, not currently included in most recent big data tools, thus make dealing with big data complex – as ‘big’ as the word sounds (Jordan, 2014).

The issues that might face big data are mostly related to complexity. The way of understanding big data and the usage of unstructured formats, such as text and video, are at issue (Mcdonnell, 2014). The first step to avoiding serious consequences of big data failures is having expert employees and managers who can handle business flow and justify the costs because it is a challenge for SMEs to become sophisticated decision makers (Jordan, 2014). Finding skilled people who are familiarised with data analytical tools (for instance, Cassandra and MongoDB) is not simple, especially for SMEs. However, decision makers are delaying their decisions regarding big data to wait for competitors, see the results, and learn from their mistakes (Kotorov, 2014). Another challenge regarding big data is IT security. Many SMEs are not skilled enough to work with the tools of big data. More information within the hands of SMEs means a higher chance of that information leaking or being stolen (Jordan, 2014). Information means power; therefore, more information means more power. That said the control of information sharing through the boundaries of organisations can create enmity among relationships across traditional power (Jordan, 2014). Beyond that, delivery of the most significant data captured in real time must get to the right people at the right time (Mcdonnell, 2014). Data storage is an additional issue, as is how to analyse data according to their size and SMEs’ computing abilities (Mcdonnell, 2014).

VIII. CONCLUSION

Big data for SMEs can create truly valuable opportunities. When using the right techniques, big data will increase SMEs’ outcomes. Considering B2B is an important aspect in addition to social media. Another important aspect are the Vs. Velocity and Variety is a very valuable V’s for SMEs and may have many dimensions and results when are applied with SMEs. But without the right vision and understanding of those challenges, the risk of turning big data from something positive and useful will be reversed – becoming a huge liability for an
organisation. In other words, using big data is a powerful tool for those who have the ability to use it properly (Jordan, 2014).

IX. REFERENCES


Big Data And It’s Value To SME Restaurants

Awfa G S Al-Adawi
School of Computing and Mathematics
University of Derby
Derby, UK
a.aladawi1@unimail.derby.ac.uk

Abstract— Large data sets are derived from the additional information that come as a result of the analysis of single and the large sets that are linked to a big data. Big data has positive impacts to small and medium enterprise restaurants and thus, should be used in the most appropriate situations as possible. Therefore, it is essential to understand the definition of big data, the four V’s of the big data and their meanings, history, opportunities, challenges, benefits and the recommendations that are applicable in a restaurant business setting in relation to handling big data.

Index Terms— Big data, SME’s, restaurant, four V’s of big data, big data governance.

I. INTRODUCTION

Big data can be very difficult to work with spatially for SME’s restaurant and may require to be broken down into simpler parts, so that interpretation is made easier in the small medium enterprises like restaurants. The main challenges that are experienced in managing big data include; the storage of data, capturing the data, sharing, searching, visualization, analysis and transfer of the data to the required personnel like restaurant managers (Zikopoulos 2012). Large data sets emerge from derived additional information.

A. What is big data?

Therefore, big data can be defined to be a situation whereby data sets like the reviews from clients collected, are complex and voluminous to an extent that a manager finds it difficult to process through tools like on hand database and other traditional applications for processing of data (Mayer-schönberger & Cukier 2013).

SAS define big data as a general term to define the speedily growing and accessibility to data, for both structured and unstructured data. In addition, big data could be as significant to both society and business as the Internet is today (Sas.com. 2013).

B. The four V’s of Big Data and what is meant by each one (As IBM)

As per IBM, the four V’s of big data include the volume, veracity, variety and the velocity. The volume is mainly meant to show the scale of the given data. The volume can be classified in terms of terabyte and petabyte to show the scale rate of the big data (Ohlhorst 2013). Variety is used to indicate the different types of data that can be obtained in a big data. A variety of data can either be structured or unstructured. Structured data enables easier management as compared to unstructured. The velocity of big data refers to the rate at which the data can be analysed in terms of the streaming data while veracity is used to show uncertainty evidence in a big data. All this are applicable in a restaurant setting for example, in financials and customer reviews (Davis & Patterson 2012). However today Big Data have more than that four V’s, It can reach around 12 V’s “Volume (size), Velocity (speed), Variety (sources/format), Variability (temporal?), Value (to whom?), Veracity (truth), Validity (applicable), Volatility (temporal), Verbosity (text), Vulnerability (security), Verification (trust?), Visualization (presentation)” (Self 2014).

C. History of Big Data

Big data as per historical terms was viewed as ‘information explosion’ in the year 1941. The term was coined from the fact that, American libraries kept growing after every sixteenth year. For example, Fremont Rider a librarian in Wesleyan University forecasted that 2 billion worth of information could fill 6000 miles of shelves in catalogues (Berman 2013). This resulted into new methods that could be used to store the large data that had been forecasted and also, kept growing is volume. Technology storage of data came into existence in order to control the large data, hence the name big data (Davenport 2014). Big data started to be known in early 1990’s at the same time it was very good time for the big data analyses computer program to found as the big company see that long back and start to find a way to work or use that big data for them side (figure 1).

Fig. 1. The time line of the BIG DATA
II. OPPORTUNITIES AND CHALLENGES FOR BIG DATA TO IMPROVE RESTAURANT

In a restaurant, big data can be used in improving the quality of service provided and supplies due to the transparency nature that it offers (Franks 2012). This is because, with big data in a restaurant, prediction, inconsistence performance, availability of specific items required and applicable approach can be determined. Through this, a restaurant can improve its performance in terms of goods and service provision (Schmarzo 2013).

However, the micro processing of big data may give assumptions that have been deduced from mathematical properties with little or no reflection of the real situation at micro processing level. All the data has to be contextualized in terms of political, social and economic settings which maybe a misplaced priority in a restaurant setting (Mohanty, Jagadeesh & Srivatsa 2013). There is bound to be biases towards a particular service or provision of goods in a restaurant that can be rectified through the use of the four V’s (Minelli, Chambers & Dhiraj 2013).

A. Benefits for restaurant to use Big Data

The main benefits for restaurants to use big data are attributed to the fact that, through big data, a restaurant can try to outperform one of its main competitors in the industry. This can be achieved from checking the reviews meant for other related industries (Chen 2014). For most of the restaurants, influencing of strategies that are data based is the key to innovation. This provides a ground for competition and capturing of value (Zikopoulos & Melnyk 2013). Adoption of big data by restaurants enables an environment that ensures quality of service and good provision and good customer relations. This in turn, aids in growth of the restaurant due to different approaches that may have been given as suggestions by the clients and customers. The chances for expansion for a restaurant business are also high due to the feedbacks that big data provides in reviews (Shmueli et al. 2011).

B. Operations needed to get and analyse Big Data from restaurant

The operations required to get and analyse big data of restaurant include having the knowledge of the variety of a data so that it can be classified as unstructured or structured. Through this, the most important data can be determined (Steele et al. 2010). This is made possible through figuring the data that is best fit to be incorporated into operations of a restaurant, planning and marketing (Manning 2013).

III. GOVERNANCE OF BIG DATA

Governance of big data is an important in the organization due to the challenges associated with voluminous data. The executives are faced with very huge chunks of data relating to the sales and revenues. The large amounts of data relating to the revenues of the organizations are managed through information systems. The big data is managed through the knowledge work systems of the organizations. All the data of the organization is input into the knowledge work management systems of the organization. This knowledge information helps in the organization and analysis of such data so as to facilitate efficient decision making in the organization. The large amounts of data are broken down into meaningful forms. The system also helps in analysing the data such that the executives of the organization can make decisions relating to the critical issues and concerns of the organization. The systems also help in eliminating the irrelevant data and giving emphasis and focus on the relevant facts (Mohanty, 2013:18). The knowledge management systems are of the essence in ensuring that executives and the overall management of the organization can make decisions without incurring high costs and expenses by concentrating on the relevant aspects of the data of the organization.

A. Privacy and big data

Privacy and big data is an issue of concern in the organization. Big data is prone to attacks and manipulation both by the employees and the external parties. For example the big data of the restaurant may be misappropriated or manipulated leading to poor decision-making as well as losses in the restaurant. The data relating to the payroll may be manipulated by inclusion of the ghost workers in the payroll system of the restaurant. Manipulation of data may lead to very huge losses to the restaurant if the data is not monitored and protected in a proper manner. The restaurant may put the necessary measures in place in order to ensure that privacy of the large data of the restaurant is secured and safe. The use of passwords and encryption of data will help the restaurant to prevent the manipulation of the systems and data of the restaurant by external or unauthorized parties. The installation of both physical and logical security measures is an important aspect in ensuring privacy and security of the information systems of small, medium enterprises. The experts should also be trained in order to ensure that the information systems of such enterprises are monitored regularly and closely so as to ensure privacy and safety of big data of the restaurant (Fearon, 2013:28).

B. Social network and big data

The social network may expose the big data of the small and medium enterprises to risks of viruses and worms. The viruses and worms may destroy or manipulate the data of the organizations leading terrible problems in using such data. The data may become meaningless and useless due to the distortion by the viruses and worms in the social network. Hackers may also use the social network to access the big data of the company leading to loss of the confidential information of the organizations. The competitors may also access the strategies of the company hindering the progress of the organization due to such leakage of such business secrets. The social network is vulnerable to malicious parties that may manipulate the data of the company leading to a bad public image of the concerned organization on the internet or the whole social network. The corporate values of the organization concerned may be deteriorated due to such leakage of information on the social networks. The internet may also lead to the aspects of
impersonation that may also deteriorate the goodwill of the organization concerned.

**IV. MANAGING BIG DATA**

The management of big data is an important aspect of concern in order to facilitate smooth and efficient operations of such organizations. Management information systems are of the essence in ensuring proper and effective management of such data. The data is stored in databases that further analyse such data into blocks and small meaningful chunks. The small groups of data are useful in making specific decisions pertaining to various aspects and departments of the organization. The data may be analysed such that the critical issues like revenues, supplies and promotional activities can be addressed fully by the organization. The information systems help in breaking the data such that accurate and simple reports are generated that the executives can use to make realistic decisions relating to the various departments and critical issues of the organizations. The big data can also be managed such that each department or level of the organization is assigned specific tasks and activities based on the data of the organization concerned. The managers of the organization in each department or level of the organization concerned, handles the data in order to make meaningful results at each level of the organization. The breaking down of data helps in facilitating faster decision making in the organization.

**A. Restaurants need to reduce waste in the food**

Restaurants most of the time makes more than what they sell, this food sometimes saved for the next day or go to the trash. Reducing the cost and effort require to make all that wasted food is important to increase profit. Some restaurant will cut cost and reduce the quantity of the produced food as an attempt to make what meet the needs of customers but they fail and lose customer because of the shortage in the food. This equation, will require managers to do a huge amount of reach, in normal circumstances the restaurant owner may be able to forecast inaccurately the customer need if he had long experience. However, he cannot forecast the amount of food that he will need in his restaurant for the next day or even for the next hours. This is due to the lack of ability to connect to too many customers and the affecting factors that combine can produce immediate conclusions of the customers need, Big Data Analysis can do that.

“Food waste at fast food restaurants usually varies depending on the type and size of the chain. Larger chains tend to have lower food waste rates ranging from 5%-7%, while smaller, local chains can have loss rates as high as 50%.” (W. Jones, 2006)

**B. Improved restaurant efficiency and BIG DATA**

IBM researchers have created a computer program that changed the food and restaurants industry that is by using the big data analytics. IBM believes that big data can improve restaurant efficiency using big data analytic computer programs not only that, IBM computer program can also generate new tasty recipes that no one came with. Within five steps between collecting and analysing the giant collection of data, such as the “of ingredients in particular recipes and the molecules and chemical compounds present in each ingredient, to human flavour preferences.” (Thusoo, 2014)

**C. Example of the data that restaurant need.**

Nowadays everyone is giving information, either by using the internet, software’s, or with mobile phones. This data combined will provide restaurants with a huge data that if analysed will lead to great profit showing the benefit of big data. Example of the different sources that may be obtained in order to get big data for the universities restaurants are:

- Timetable of students and staff, the computer program can analyse the time that the student and the staff are free and available in the university, so there is a chance that they may visit the universities restaurants. (this is as an example of Volume in big data )
- The student attendance for last year as graph will show the percentage of students that attend and can forecast when more students will come to prepare more food. (This is an example of Veracity in big data)
- Weather forecast and condition will have a direct impact on the restaurants sale; today there are many data and calculations that are being done to forecast the weather for near future, this data along with the historic weather forecast could be used by restaurants. These data and further analytics helps restaurants predict the type of customers, the food they might like and the number of expected customers. (Kühn, 2012). (This is an example of Velocity in big data)
- Events schedule and activities calendar which are happening in many different levels such as Country, City, University, society and community, Student groups as well as religious occasions.

(Moreover, all of the above points can be counted as the Verity in big data)

Computer program have the ability to analyse this information and many other information from different resource to find the relation and what is item need to produce.

**V. RECOMMENDATIONS**

Big data should be handled in the best way possible. The management department in an SME restaurant should classify data in terms of the volume of the data, variety, velocity and veracity for easier access and analysis (Isson & Harriott 2013). Through considering this, it is possible to obtain the required data in relation to a specific setting as a restaurant (Christian 2011). Big data is voluminous in nature and, therefore, there should be measures that are considered by most restaurants and small and medium enterprises in business settings when analysing financial and feedback.

**VI. CONCLUSION**

For a business to be successful, big data has to be analysed via effective and appropriate measures as the results are used to forecast the current views into improving the future conditions. The challenges encountered in big data management can be
minimized through classifying the data according to their varieties.

VII. REFERENCE


Big Data Analytics
In A Qatari Travel Agency

Hassan Al-Emadi
School of Computing and Math
University of Derby
Derby, UK
h.alemadi1@unimail.ac.uk

Abstract—The purpose of this research paper is to identify “Big Data” technology and demonstrate the opportunities, challenges, benefits and operations of Big Data. In addition, it will do so in the context of Small and Medium Enterprises (SMEs).

IndexTerms—SMEs, Big Data, Challenges, Benefits, Travel Agency (keywords)

I. INTRODUCTION

What is Big Data? When someone hears this terminology, the first thing that may come to mind is huge, massive, and really big data containing a lot of information. Big Data is defined as an extremely huge amount of data that cannot be managed and analyzed with the traditional way of processing the data. As a result it is difficult to get the value of that huge amount of data, and the most difficulty will be in its acquisition, analytics, searching, storage and sharing. As data grows each year, analytics have arrived in the IT field, led by IBM and SAS Inc, and large brands are trying to benefit from Big Data Technology (Ohlhorst, 2013).

II. BIG DATA ANALYTICS

Big Data Analytics is a big part of the job for companies that want to benefit from the huge amount of data and information. Here it is necessary to use the most advanced techniques to analyze and operate Big Data, focusing on the most V’s for the company (Russom, 2011).

III. THE THREE V’S OF BIG DATA

When the people talk about Big Data they mention the size of the data and how it is so big. But there are other aspects that need to be considered, such as the V’s of Big Data: Volume, Variety, and Velocity. The following sections will discuss more about these three V’s and how they matter (Russom, 2011).

A. Volume

It is very clear that volume of data is most characteristic of Big Data. These data could be pictures, videos, voice files, documents, feedback and much more. They could take up huge storage capacity and could be stored for a long time as it is in the USA. Some businesses do not worry about the quantity or size of the data in terabytes, but it matters when it comes to the number of records and information (Russom, 2011).

B. Variety

The biggest reason that makes the attribute of Big Data really huge and Big Data is that the information and data are imported for different resources and different types. The major source is the Web, including social media, general websites and logs. The variety of the data is important but makes it difficult to gather this information, especially if the company needs information from different organisations such as the government, airports, ministries and statistical organisations. Hence HADOOP is one of the favourite and recommended systems that can take care of both a large volume and wide variety of data (Russom, 2011).

C. Velocity

Velocity is a very sensitive V because it is about the time and sequences of the data. The company should get the benefits from Big Data and analyse it in time so that the data does not become useless. As a result, the data imported should be gathered as soon as possible and provided the best analysis that can be used to improve business (Russom, 2011).

IV. ANALYSIS CATEGORIES

Analysis of Big Data relies on a combination of several technologies and concepts. The following are some of the strategies used:
A. Traditional Business Intelligence

A combination of technologies and applications that can gather, store, analyse and allow access to data, usually using database software (Ohlhorst, 2013).

B. Data Mining

Data mining is a process for a large amount of data with special techniques and tolls, which can export a summary of useful information, and it is usually used for archived data (Ohlhorst, 2013).

C. Statistical Application

This kind of technology uses a set of statistics and algorithms to look to the data, and it is used for survey data (Ohlhorst, 2013).

D. Data Modelling

It is used for “what-if” scenarios using algorithms for multiple data. The model may change based on the information available to the algorithms (Ohlhorst, 2013).

V. HOW BIG DATA IS DIFFERENT

These days many people in the business sector are mentioning and talking about the new terminology and technology of Big Data. But each business is explaining and understanding this phenomenon in different ways. People in the business sector always cooperate with the technology sector to find better business solutions that can prove more reliable in business, which will lead to increased profits and business expansion. Everyone knows that many companies these days have a huge amount of data, even small and medium enterprises (SMEs). Many of these companies have no structure to manage or analyse such data. Data can come from user feedback, social media, government organisations or personal data like photos, videos, voice and biometric data. But Big Data is also much more than that. IT companies are now using real-time data to analyse and establish new products, services or changes in usage (Davenport et al., 2012).

To clarify how Big Data is different, IBM said such data typically has four attributes, the four V’s of volume (terabytes to petabytes of data); variety (data in many forms, structured, unstructured, text and multimedia); velocity (analysis of streaming data to enable decisions within fractions of a second); and verity (managing the reliability and predictability of inherently imprecise data types) (Russom, 2011).

VI. WHY BIG DATA MATTERS

Even knowing the definition of Big Data and Big Data analytics will not identify the needs and benefits of this technology. These days, meeting customers’ requirements is most important for enterprises, for finding that new technology or new concept to communicate and identify customers’ needs is one of the top priorities for companies. Some companies have had successes in analysing what customers need and what they will buy in the coming days. For example, e-commerce websites use all data available to predict what customers are looking for; if a customer is looking for a jacket in winter, they can also be offered tea or coffee products. By gathering information from customers to provide new offers, these kinds of strategies can improve business growth (Ohlhorst, 2013).

VII. WHAT CAN BIG DATA DO FOR YOU?

A travel agency is the same as some other areas of business. They are using Big Data headlines to identify and improve the business and customers’ needs, as well as facing the challenges, benefits and impact. The use of the Web, smart phones, cookies, email, sensors, different applications, social media and e-payment creates a huge volume and variety of data in the Internet world. Here the terminology of Big Data become observe to the people and companies, that happens because of the technology that can happen with mining, store, save, analyze and export when the user need it (businesstravelnews.com, 2013). “Every day, we create 2.5 quintillion bytes of data — so much that 90 percent of the data in the world today has been created in the last two years alone,” according to an IBM briefing (businesstravelnews.com, 2013).

Figure 2 shows that 34% of managers and buyers have started to talk about Big Data and ways to use it. On the other hand, there are 25% of managers and buyers who do not know what they don’t know about Big Data (businesstravelnews.com, 2013).

Data from social media and customers feedbacks, either on the Internet such as tripvisitor.com or the feedback provided by customers to the travel agency; data imported from government organisations such as the Ministry of Statistics and airports; data from customers such as a customer profile — these are all resources that can be the best part of a success story of any travel company. In addition to running analytics and mining tolls, these data need new technologies such as that provided by IBM (businesstravelnews.com, 2013).

VIII. SMEs AND BIG DATA

A. How we can help SMEs to benefit from Big Data

Big Data is always available, but the question is how big it should or can be for SMEs and how that can help these
companies if they know what they have and what Big Data is. Many SMEs don’t know actually what big data is as technology and how they can use it, because Big Data is a combination of multiple technologies at the same time (Macinnes, 2014).

B. Apply Big Data on a Small Scale

According to Campbell Williams group and marketing director at Six Degrees Group; “There’s a simple reason why there’s no market in small and medium-sized enterprises for big data — they don’t have big data” he says. And it is not just SMEs, he notes: “Most large and multinational enterprises don’t have big data either” (Macinnes, 2014).

In addition, he mentions that mega-brands have a large amount of data on their customers around the world, and SMEs needs CRM (Content Management System) or ERP (Enterprise resourcing Planning) system, a good strategy for contacting customer and marketing consulting, which would be better than looking for Big Data. On the other hand, Dave Becerra, vice president of strategy and business development at Roambi, says SMEs are starting to realize that Big Data can make a difference in their business, and identifying the volume and verity of the data even if the SMEs do not own it. In addition, there is no need for SMEs such as travel agencies to own the data or produce it (Liebowitz, 2013).

There are many resources that can be valuable sources of data. For example, airports and travel agencies can cooperate to gather information on travellers and their preferred destination. Statistical organisations can provide statistics on the preferred destinations for people in different seasons or for other purposes. Governments can help with records of travellers’ who travelling for education, treatment, tourism and the length of the trip. Also, social media can be the most effective source of data here, such as identifying traveller satisfaction, and can help to complete the data from the previous sources (Macinnes, 2014).

C. Big Benefits for Travel Agents

A side from Big Data that can be collected from different sources, SEMs should get the attention from it is data. Travel agents have at least two types of data they may not realise. First is the large number of user profiles and booking history for each customer that has been already created. Second is the data produced by different travel companies worldwide each day, such as itinerary details and booking searches. Travelport.com processes a massive amount of data each day, and can benefit from analysing the real-time data and exporting important data that can add a great value to the business. With these powerful data resources, a travel agency can analyse, identify and discover new destinations that meet traveller requirements and interests customise advertisements for the customer’s travel style, and set offers based on popular destinations, ages or purpose of travel. That can lead to extrapolating the future behaviour of customers and planning a strategy for the next steps (Travelport.com, 2014).

D. Specific Benefits for Travel Agents

Some other specific benefits will be observed such as what customers need and how best to provide support for client issues. Also, the travel agency can offer new offers and products by knowing what kind of activities tourists would like in certain places. The agency may discover new places it did not know about, possibly from social media and people sharing their favourite journeys (Digital Tourism Think Tank, 2013).

IX. HOW BIG ARE THE CHALLENGES?

As we know the amount of the data is too large for traditional database systems to process, manage and show results in a suitable time before the data become worthless. Enhance, the technology cute long term of development and researching, and now these super-powered IT solutions are not only for big organisations, but also for SMEs like travel agencies (Travelport.com, 2014).

But SMEs could face some challenges. First of all, using and implementing new technology could cause some management issues and issues with employees. Secondly, the cost of building Big Data infrastructure could cause some financial issues for some SMEs (cra.org, n.d.).

Ensuring sustainable resources of data and information is also a challenge, because the data itself is the most important characteristic of Big Data. If one resource goes down, there will not be a good value in the rest of the data. In addition, providing and reaching the speed needed to process and analyse the data with good exported results requires understanding the input and output data with powerful equipment in order to help humans to benefit from it (sas.com, 2014).

X. ETHICS AND PRIVACY

It is common to find some privacy and ethics issues in any technology around the world, especially if it involves direct communication and interaction with human users. Big data analytics for Qatari travel agency will face few issues of people using multimedia files such as photos, videos and voice, especially because Gulf States are culturally traditional countries that have different customs and traditions. Hence, SMEs should be careful when dealing with these data and apply terms and conditions for privacy purposes, as with the grocery store which used customers’ chosen items and behaviour to identify the next acceptable and item for purchase. That can be done, for example, by noticing users’ “likes” on specific pictures or videos for specific destinations that maybe the future destination and interesting to the user (Duhigg, 2012).

XI. CONCLUSION AND RECOMMENDATION

In conclusion, Big Data is a very powerful technology, and data and analytics are needed to reap the benefit of this technology. In this research it is recommended that any company wanting to get this technology plan a strategy and study the company case, while also ensuring the data resource. In addition, the company should take care and focus on some bullet points. For example, identify the target goal of the
project and how this technology will improve the weaknesses of their business. Analyse your data first then ask for more data from other resources, so as not to overload and lose the value of the data in its huge quantities. Once you have finished off these data, ask for more from different resources that provide verity of data. Finally, analyse pieces of data and focus on each sector of data. That will make the big data more beneficial to the business and will reduce the challenges and issues (Forbesmiddleeast.com, 2014).

XII. REFERENCES


Abstract—This report aims to provide a small to medium size enterprise with the information it needs to help make an informed decision about whether big data analytics (BDA) is right for them. It will focus on explaining the benefits to using BDA and also reflect on the risks and security issues associated with big data.

Index Terms—Big Data, Advanced Analytics, Security, Data Protection, Auditing

I. INTRODUCTION

The world of big data analytics (BDA) is a rapidly growing one. Prickett Morgan (2012) highlights that the big data market is expected to grow at an annual compound growth rate of 39.4% and reach a value of $16.9bn by 2015, contrast this to its value of $3.2bn in 2010 it is very clear how quickly this technology is being adopted and the effect it is having on the business environment. Interestingly he goes on to say that the big data market is growing at a rate seven times that of the whole of the rest of the IT market. This report is going to look at the benefits BDA will provide a small/medium enterprise (SME) and what security issues may arise from using BDA, so that you can make an educated decision on whether BDA is the right tool to further build up your SME.

II. FIRST OF ALL, WHAT IS BIG DATA ANALYTICS?

Of course before looking into what BDA can do for you, you need to know what it is. Prickett Morgan (2012) quotes IDC’s definition of big data as a data store that contains at least 100TB of data, and companies are storing data in memory and processing it within that memory without pushing it out to disk. IDC also say that big data is about “data in motion”, this is the capture and processing of real time, streaming data. They also go on to say that to be classed as big data, data sets have to grow by at least 60% a year; Also that the data systems are deployed on “scale-out architectures”, this basically means on clusters of servers and storage. Finally they state that systems should be “wrangling” with data in at least two different formats. “The three Vs of big data (volume, variety, and velocity) constitute a comprehensive definition, and they bust the myth that big data is only about data volume.” (Russom 2011) also supports this definition. So the term “big” in BDA doesn’t just mean you need a lot of data (Volume), it also needs to be of different formats (Variety) and it needs to be moving or in real time (Velocity). These are the factors that makes BDA a powerful tool.

A. How can big data analytics help your SME?

Big data analytics has the potential to aid your SME in taking its next steps towards intelligent business decision making. “Using big data enables managers to decide on the basis of evidence rather than intuition. For that reason it has the potential to revolutionise management.” (McAfee & Brynjolfson 2012). With more effective management decisions achieved through the use of big data analytics your SME will increase its revenue, providing the right decisions are made. Bromhead (2014) states that “Big Data Analytics, including Predictive Analysis, can help companies identify opportunities to boost revenue, service quality, brand image, and even job creation.”

B. What market intelligence does BDA open the door to?

Huster (2005 cited in Trim & Lee 2008) suggests that “the ability to fully understand, analyze, and assess the internal and external environment related to a company’s customers, competitors, markets, and industry to enhance the tactical and strategic decision-making process”. This evidence clearly points out that the more an SME can find out about the factors highlighted by Huster the better an SME will perform. Mayer-Schönberger & Cukier (2013) refer to a case study conducted by Google where they were able to predict the locations of outbreaks of the flu virus based on internet searches. This was done by using the vast amounts of data that Google collects every hour and analyzing it so that these predictions, based on the evidence BDA had provided them with, could be made. Using this method an SME could predict customer behavior using real-time data and make effective business decisions to increase revenue. It is looking like BDA is a very valuable tool to be using with SAS (date unknown) stating that “big data – through high-performance analytics – could add £216 billion to the UK economy by 2017”. This raises the question of is there any competition to BDA?

C. BDA vs. Data warehousing (DW)

A data warehouse is defined as by Inmon (1993 cited by Atkinson 2001, p.35) “A data warehouse is a subject-
orientated, integrated, time-variant, non-volatile collection of data in support of management’s decision making process”. Compare this to the BDA definition where the data has to be volatile, and it is this velocity that allows its users to make informed business decisions in a much shorter time scale this is supported by “One of the key differences between traditional data and Big Data is not just its size but its form. It is constant, live, and flows through an organisation, rather than be a historical lump of data that is not going anywhere” (Anon. 2013). This rapid decision making can be seen in figure.1(Vellante 2013) where the breakeven point is reached much more quickly using BDA then DW.

This clear reduction in time scale to breakeven suggests that effective business decisions can be made much more quickly using BDA and in terms of extra revenue generated could be worth $10s of millions to an SME. From this evidence it shows that BDA does still have competition from the techniques that have been used by businesses to extract business intelligence for several years now. However it does appear that BDA provides the greater reward but does it come with greater risk?

III. THE COSTS, CHALLENGES, AND RISKS ASSOCIATED WITH USING BDA

A. Consider the cost

All businesses are constantly looking for ways to improve their situation especially SMEs, where the wrong decision could lead to irreversible damage to the company, and BDA can provide an SME with the information to make the right decision and reduce the risk of catastrophic failure, however it needs to be done cost-effectively. Herodotou et al. (2011) supports this idea stating that “Timely and cost-effective analytics over “Big Data” has emerged as a key ingredient for success in many businesses, scientific and engineering disciplines, and government endeavors”. But do BDA and cost-effectiveness go hand in hand? Cicconi (2006) makes the suggestion that Technical innovation comes with high costs. Bryant et al. (2008) has tried to put a monetary value on the start-up of BDA within a company claiming that the required hardware would cost around $50 million mark and an estimated running cost of BDA is to be around $10 million per annum. Refer back to figure 1, assume that the initial loss is the cost of the software only. Adding the initial cost of the hardware, which most SMEs will need to acquire before starting to use BDA, it begins to paint a new picture. The graph shows that the cumulative cash flow doesn’t break the $50 million mark until around 20 months after the initial investment therefore an SME that has made this investment would not see a return on their investment until around 20-24 months down the line. However this value was assigned to the technology in 2008 and Moore’s law states (Moore 1965 cited by Rouse 2005) that the number of micro components that could be placed in a microchip at lowest manufacturing cost was doubling every year. More recently this has changed to every eighteen months. What this means is that the technology that would have cost $50 million in 2008 would only cost around the $3.125 million mark today, providing the processing power remained the same. Or you could have a substantially more powerful IT infrastructure in place for your $50 million.

B. What are the security issues associated with big data?

As with any new technology there will be new security threats and risks that need to be assessed so that effective governance strategies can be put into place to prevent these risks turning into a SMEs worst nightmare. One security issue linked to BDA is caused by how complex this analytical process is, due to its very nature, the variety and velocity of the data, BDA proves increasingly difficult to keep the whole system secure. This is supported by Audenstad (2005 cited by Werlinger et al. 2009) who suggests that a reason one hundred per cent security coverage isn’t achieved is because of the complexity of technology. Coming back to this idea of complexity. MongoDB inc (2014) states the typical company dealing with big data studies many data types from text to video files in order to build a complete picture of their business environment. Combine this with data being stored in clusters of servers rather than one centralized server as with traditional databases the system as a whole becomes very complex. They go on to mention that Facebook ingests over 500 terabytes of data a day which only highlights the fact that BDA is a technology that will be difficult to keep completely secure, you only have to look at the what happened to Target between 27th November and 15th December 2013 when for this entire period cyber criminals were able to gain access to and export over 40 million customer accounts including credit/debit card numbers and the corresponding security codes for each card. This clearly highlights the need for tight security and regular audits to ensure no breaches like this ever occur because to an SME this could be crippling. Incidentally Skariachan & Wahba (2014) report that retail companies are only spending four per cent, banks are spending five and a half per cent, and healthcare companies are spending 5.6 per cent of their respective Technology budgets on security which is one reason why so many governance failures are occurring in relation to IT security. To effectively manage big data and keep it secure, enough should be spent to maintain high levels of security and consistent (maybe even constant) auditing as a minimum. Now this may seem overkill to some SMEs but with the rise in
economic crime, Price Waterhouse Cooper (2014) report that 37 per cent of companies report this type of crime (that’s a three per cent increase from the previous year), they then go on to say that of these reported cases 24 per cent of the companies are victims of cybercrime. This evidence clearly outlines the need for effective security management, which may come at a higher cost, but offset this cost against how much an SME will lose in compensation claims and law suits should the worst happen.

C. Keeping track of your data, is it even possible with big data?

“Today’s business requires that a company’s data are managed in a centralized manner”. (Silvola et al. 2011) which sounds logical (but is it possible with big data?) the very definition of big data is that it is volatile, it’s constantly on the move, growing in real time and adapting to changes in the business environment, which is good for the decision maker that wants to remain on top of his or her market; the problem arises when the data is not in a centralized location, the data still needs protecting from attack especially when the data is of a sensitive nature. To ensure that BDA is working efficiently for a SME a balance between the freedom the data has to grow and the security strategies put in place to manage the data needs to be met. After all spending that bit more to keep tabs on all of your data should prove to be the right decision, Snow (2008); Batini et al. (2009); Redman (2001) (cited in Silvola et al. 2011) justify this claim saying that incorrect or mismanaged data costs the retail industry alone $40 billion annually in missed opportunities, failed deliveries, invoicing issues and mistakes. Effective data management and governance strategies should prevent these missed revenue opportunities and keep the data secure from any malicious activity.

IV. WHAT ARE AN SMES RESPONSIBILITIES WHEN USING BIG DATA?

To quote Franklin D. Roosevelt “great power involves great responsibility”. And big data generates a lot of power for those who can analyze it effectively. With this though comes the responsibility to ensure the data is managed properly and protected against sources of harm, whether that be human threats or non-human such as virus etc.

A. What Does UK law say?

All companies registered in the UK must adhere to the DATA PROTECTION ACT(1998) which states in Schedule 1, part 1, section 7 that “Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data”(Great Britain, Data protection act 1998). Therefore any data stores that fall under a SMEs control/ownership (also known as provenance) need to be protected by adequate security practices to prevent a breach of the law that could lead to substantial monetary penalties and sanctions. The act also states in Schedule 1, part 1, section 8 that “Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data” (Great Britain, Data protection act 1998). Due to big data is so volatile extra efforts need to be put in place so that a SME doesn’t fall foul of this section of the act; especially when dealing with third parties outside the EU or if a decision is made to outsource some data to provide an extra revenue stream.

B. What steps should a SME take to ensure the security of its big data?

Keeping its data safe is vital to any company, and this does not change with big data. If anything it becomes all the more important. The International Organization of Standardization (ISO) states in chapter 5 section 5.1 sub-section 5.1.1 “at the highest level, organizations should define an “information security policy” which is approved by management and which sets out the organization’s approach to managing its information security objective”. (BS ISO 27002 2013). The big data that a SME using BDA keeps becomes that company’s asset and chapter 9 section 9.1 sub-section 9.1.1 suggests that “asset owners should determine appropriate access control rules, access rights and restrictions” (BS ISO 27002 2013). With the volatile nature of big data, information transfer policies should be in place, this is supported in chapter 13 section 13.2 sub-section 13.2.1 which advises that “formal transfer policies, procedures and controls should be in place to protect the transfer of information through the use of all types of communication facilities” (BS ISO 27002 2013). Following these guidelines, an effective governance strategy can be implemented to keep data secure and prevent malicious (maybe even damaging) behavior. Finally a SME should consider Disaster recovery Hawkins et al. (2000) support this idea by saying “The migration from centralized mainframe computers to distributed client/server systems has created a concern on data security. If a disaster occurs to the organization that destroys a server or the entire network, a company may not be able to recover from the loss. Developing an effective disaster recovery plan will help an organization protect them from data loss”. Since Hawkins et al. wrote this disaster recovery has become known as ‘business continuity’. A SME should have a business continuity plan in effect and guidelines for this process can be found in the document ‘BS 25999’ available on the British Standards website.

C. Is there any protection available to the SME if something should go wrong?

In short the answer is yes, protection of your data is available. The catch is you have to give your data a monetary value. Vladimirov (2014) says that only one in ten companies assigns a monetary value to its data which he admits to being difficult to do due to there not being a tried and tested method in place. He does however highlight that IBM a few potential strategies by measuring the data’s volume, variety and velocity (known as the customer lifetime value) a company can determine how much revenue its data is responsible for and give it a monetary value. Once this is done the data can be insured.
V. CONCLUSION & RECOMMENDATION

To summarize big data clearly has advantages over other data analysis techniques. Allowing companies to make real time business decisions being just one of them as well as the potential earning power it seems to provide. However as with all emerging technologies it has its risks and security issues, the incident with Target in late 2013 being one of the most high profile security failures with regards to big data to date, meaning that substantial running costs are going to have to be endured to ensure the security of the data being processed/stored/outsources.

From the evidence gathered in this report big data looks like a winner but only if the company has an effective governance strategy in place to ensure its security and efficiency. The start-up costs need to be addressed as well, an SME looking to implement a big data analytics venture of its own must be in a position to cover the costs over a substantial period of time before a return on the investment is seen, do refer back to figure 1 in chapter 1 once a return is seen it is quite a large one. To that end, with the right strategy and controls in place big data could quite possibly make the reward worth the risk.

VI. REFERENCES


Abstract—The quantity of data today has been increasing at a rate that has never been imagined before. Small & Medium Enterprises (SMEs) capture millions of kilobytes of data about their suppliers, clients, daily operations and multiple sensors embedded in vehicles, mobile phones and others. Big data goes beyond storage of huge amounts of data. It also involves giving capacity to ensure better decisions are made so that managements can act accurately and on time. SMEs capture data with superior frequency and granularity, collecting all client transactions while attaching personal information such as credit card number and monitoring consumer behavior. The greater the collection of data, the greater the necessity for additional storage in addition to analytical applications to ensure all the collected data is stored for future access and reports to support decision making. This paper seeks to discuss big data and its influence on SMEs.

Index Terms—SMEs, Big Data Technology, Cloud technology.

I. INTRODUCTION

Big data refers to a collection of complex data sets. Large data sets that process its using conventional data processing systems or readily available database management implementations are demanding (White, 2012; Big Data Explained n.d.; Davis, 2012). Franks (2012) asserts that big data results from extra information derived from several analyses of a collection of interrelated data.

Today, there is unexpected growth of data sets as they are being collected by new and emerging technologies such as IP cameras, software logs, Radio-Frequency Identification (RFID) detectors, ubiquitous mobile devices, remote sensing, and aerial sensory tools (EMC, 2014). The challenge faced by SMEs is deciding on who should be responsible for owning big data proposals that span the whole organization (Data Science Series, 2012).

SMEs are finding big data demanding to work with especially using conventional relational DBMSs, visualization packages and desktop statistics (BLEUM, 2013). However, SMEs take different courses of actions to remedy this, for example, some use extreme congruent applications running on many servers (Gartner, 2014). The capabilities of an organization determine the size of data set that is labeled “big data” (Davenport, 2014). This depends on the organization’s capacity to effortlessly tackle huge data sets. Some will consider data sets to be “big data” after using terabytes while other thousands of terabytes and so on depending on their capacity to manage the data sets (Akerkar, 2013).

The emergence of increased volumes of high velocity, with enhanced verification mechanism and effective visualization in terms of presentation at low cost, coupled with a significant increase in the array of sources of data has led to the data explosion (Groenfeldt, 2011; Henschen, 2014). BLEUM research shows that, over 90 % of today’s data has been generated 2 years ago (BLEUM, 2013).

The global data storage has shifted from analogue media such as magnetic ribbons to digital media such as hard drives while greatly increasing the amount of data that can be stored on a single device or a set of devices. Today, about 2.7 Zeta bytes of data are in existence in the digital world. Since the start of the digital age, there has been a rapid growth of data. Today, there are over 123 billion gigabytes of data stored digitally on hard drives from structured and unstructured sources such as Customer Relationship Management (CRM) and social media, respectively (Gartner, 2013). SMEs have seen their data grow constantly from internal information systems and external social media thus they have also experienced big data. SMEs can gain immense benefits if proper analysis is conducted on this big data in order to obtain real-time and on-demand reports to support strategic decisions (Gartner, 2013). However, challenges such as data volume, vulnerability, visualization, complexity and big data expertise must be handled well to leverage significant benefits from big data. Gartner’s study showed that SMEs is faced with significant challenges when handling big data, including lack of big data experts, increased volumes and velocity of data among others. Big data has the potential to place an SME to a competitive edge over its rivals. In another research by McKinsey Global Institute in 2011, it is evident that big data can result to increased profit margins in Small and Medium Businesses (SMBs) by approximately 60 % (Manyika et al, 2011).

II. BIG DATA: QUESTIONS OF OPPORTUNITIES

The arrival of big data is alarming in that it is coming from numerous sources at a high variety, volume and velocity (IBM n.d.). Also, concern of big data is increasing due to its variability where the data is assessed if its temporal, its value where its assessed to whom does the big data is valuable to, veracity of the data where truth of the data is assessed, validity
of the big data where the applicability of data is assessed, volatility of the big data is assessed if it’s on temporally basis, verbosity of the data where the data is assessed on how it can be read, vulnerability where the data is determined if it can be able to maintain the clients or owners information confidentiality, integrity and availability, verification where the trust of the data is assessed, and visualization where how the big data can be presented is assessed. Optimal skills, analytics capacity and processing power are needed if at all an SME wish to extract significant opportunities from the big data sets (ISACA, 2014; Jones, 2014).

Big data has facilitated the way employees within organizations collaborate (Kuketz, 2012). Laney (2001) found that big data has led to a culture where business management, as well as IT personnel, must work together to achieve benefits from every data set. Big data has brought about insights that enable a business to make better decisions including optimizing operations, strengthening customer engagement, avoiding fraud and threats, presenting new revenue sources and improving public relations to enhance the business image (Maclelnnes, 2013). SMEs possess big data and it comes with a lot of opportunities in the form of insights to make better decisions and take the best course of action when presented with a variety of challenging options to choose from.

The escalating insights require significantly new approaches to practices, tools and architecture (Mayer-Schonberger & Cuckier, 2013; Matteson, 2013). Big data provides a resource for achieving competitive advantage, empowerment to better decision-making and increasing value of data (McGuire, Manyika & Chui, 2012). SMEs can place themselves at a competitive edge through proper use of big data (Rubin, 2013). They need to use concrete analytics and exploit the value in every type of data. To achieve competitive advantage, SMEs must use the relevant infrastructure to effectively manage increasingly growing volumes of data both in motion and at rest and safeguard data security and privacy.

The infrastructure must exploit real-time information moving in the SME as well as have optimized analytics to dynamically respond to business processes using big data so as to realize improved economics and better agility (Rubin, 2013). Efficient analytics must be run close to business data while in motion or at rest to realize efficiencies and economies. But for the data to be stored, the infrastructure must have a defensive strategy that minimizes the legal issues, storage rate and other risks. Greatest value can be obtained from big data upon acquisition of new skills to fully exploit. SMEs must develop new roles focused on primary challenges, in addition to creating new models, to benefit greatly from big data.

III. BIG DATA: CHALLENGES

Big data provides significant value but still presents significant challenges. SMEs must be practical enough to safeguard privacy, governance and security to guarantee protection of all insights and data (Sestini, 2012). Mann (2012) and Peck (2013) assert that privacy of data is greatly threatened by big data. In their research, Michael & Miller (2013) note that operations on big data are faced with several challenges including acquisition, storage, verification, visualization sharing, search, transfer, visualization, business informatics and analysis among others in that it takes the organization intolerable amount of time to carry out the tasks.

SMEs should, therefore, install an infrastructure that comprises solid practices and policies to protect their brand and reputation. The infrastructure should have solid security policies to safeguard the SME against threats so as to achieve competitive advantage. Shortage of big data talent and people with profound expertise in analytics is a constraint to extracting opportunities and benefits from big data (ISACA, 2014). Furthermore, it is difficult to create the required talent and expertise in time to tackle the increasingly generated data. Most SMEs are budget constrained to secure experienced big data personnel, therefore, they may not fully enjoy the value presented by big data.

Big data leads to ethical challenges, for example, businesses use big data to monitor the workforce so as to increase productivity. Tracking employees continuously introduces an oversight and surveillance that can destroy employees’ mood (Sestini, 2012; Peck, 2013). How will the data stored be maintained and at what cost?

Most SMEs are not equipped both in their expertise and tools to manage either the huge amount of data or the velocity of data generation (McDonnell, 2011). Means of data storage are cheap but parsing, transforming, scrapping and formatting create a major challenge concerning the lack of structure and huge variety of data (McAfee & Brynjolfsson, 2012; Lohr, 2012).

IV. BIG DATA: BENEFITS

Big data has several benefits in that it presents transformational prospects to extract value (Manyika et al, 2011). SMEs are offered a platform to exploit big data residing in their databases and that in motion from which they can gain tremendous benefits. Based on the big data, SMEs have advanced marketing abilities that evolve based on the data analysis. From the large pool of data, which can be easily accessed, processed and analyzed, SMEs are empowered to analyze suppliers and customer trends, as well as compare quality and prices with those of competitors. By availing big data to all departments, SMEs can search and process data more quickly. Integrating big data from all business units enables concurrent operations and potentially reduces marketing time, in addition to improving quality (Ohlhorst, 2012).

As more data is stored digitally, SMEs have the capacity to capture more detailed and accurate performance data in real time ranging from personnel absent or sick days, product inventories, high demand goods and other important information. SME strategic managers, therefore, have the capability to make sound decisions and on time based on this information. Unnecessary costs and time wastage are successfully avoided since potential risks can be mitigated on time (Kuketz, 2012).

Big data generates high financial value and efficiency in businesses including SMEs. Manyika et al. (2011) note that US
Big data provides SMEs to turn their data sets into revenue sources. Big data allows SMEs to make better pricing. Supermarkets and other stores to identify significant client behavior patterns, relationships to client loyalty and showcase trends are using data collected from social networks such as Facebook. Responding to customers demand is therefore accurate and fast due to consolidation of customer data from multiple sources and subsequent presentation of the data in a single rapid view (Ohlhorst, 2012).

V. BIG DATA: OPERATIONS

Big data and related analytics are generally viewed as the realm of the big corporations. Therefore, many SME management teams end up ignoring the several opportunities these implementations can present to their small and medium businesses. Unfortunately, this oversight can render SMEs unfit leading to significant loses out to business rivals who are efficiently using big data to improve performance, as well as gain new and important insights (Simons, 2013).

SMEs commonly possess inadequate resources characterized by small budgets but they have other strengths: typically, they have highly flexible IT resources, with small number of legacy computer and information system issues and separate databases, and can change practices more quickly. With several specialized software systems and services in the market, SMEs can integrate existing internal business data with externally available data to discover novel insights. However, traditional big data tools from key vendors are too expensive to SMEs but there exist many cloud-based and online systems, for example, Kaggle, Tableau and Google Analytics, available to help SMEs analyze data effectively. These tools enable SMEs access powerful techniques to realize meaning of their business performance (Simons, 2013).

VI. BIG DATA: CHALLENGES SOLUTIONS

Big data challenges can be solved if the SME’s moving their data into a secure cloud where they can enjoy the benefits that the cloud computing solution offers (Gartner, 2014). By taking their big data into the cloud the SME’s overcomes challenges such as storage, sharing, search, transfer, visualization, business informatics and analysis, since cloud computing provides unlimited storage capabilities, effective data access and sharing capabilities, and updated informatics and analysis applications through its various architecture mode (Nielsen, 2011). Some of the models in which SME’s can access their big data on the cloud are through Software-as-a-Service (SaaS), where the SME’s will access management, analyzing and processing software as services, Platform-as-a-Service (PaaS) where the SME’s can lease the clouds operating systems, database management systems and other programming languages which they use to analyze or process their business processes, and Infrastructure-as-a-Service (IaaS) where the SME’s leases the server to store their big data and analytics applications(Nielsen, 2011). According to Jamsa (2012), some of the benefits that SME’s can gain by moving their data into the cloud as a solution to the challenges that they face are the SME’s will enjoy large and efficient storage capacities for their data which will accommodate the growth of their data without incurring additional costs, they will enjoy the benefits where they will be able to analyze large amounts of data using the clouds provider analyzing tools without incurring costs in purchasing or updating the analyzing applications, since the clouds will allow the SME’s to store large data, the SME’s will be able to make quality, stable and reliable analytics since they can generate correlation and reveal hidden patterns in every data that they ever had, and since most cloud providers are connected with redundant networks where they offer their clients with 99.9999% availability of their data. Also, the will have scalability advantage over other SME’s who have not taken their big data into the clouds since they can expand by establishing other regional offices in different geographical locations since they can access their data and respective processing analytics applications without added costs.

VII. CONCLUSION

Big data offers opportunities and challenges when implemented in the SME’s where it’s the mandate of the SME managers need to understand the opportunities and threats presented by big data and must assess several elements so as to align their data strategy to the necessary mechanisms to obtain relevant opportunities from big data. SMEs must be proactive and creative in the process of determining the datasets they can integrate to realize the value while addressing several issues including security, privacy, big data talent and expertise and other challenges so that the benefits of big data can overrule the challenges presented.

VIII. REFERENCES


Opportunist For Supermarkets
Big Data Application In Supermarkets

Adel Alsooj
School of Business low and computing
University of Derby
Derby, United Kingdom
A.Alsooj1@unimail.derby.ac.uk

Abstract—The use of big data has been increasing all over the world and hence companies are trying to make maximum benefit out of the same. Supermarkets having large number of customers, are excellent sources of big data. But it is found that several supermarkets in the world, especially the ones in developing countries do not use big data and do not leverage its capabilities. While supermarkets like Tesco and Sainsbury are making complete use of its existing customer base, supermarkets like Al- Meera still have a long way to go. It has thus been shown that collection and use of big data by Al- Meera can help it increase its profits and also increase customer satisfaction and loyalty.

IndexTerms—Big Data, Supermarkets, Al- Meera Supermarket, Qatar, Profits, Customer Management

I. INTRODUCTION

Information technology (IT) has always played a major role in organizations all across the world. The organizations not only use IT services and technologies for managing their basic operations but also keep improving them for enhancing their overall services and offerings to the consumers. The use of analytics that is completely driven by the collected data in an organization significantly assists companies in evaluating their performances and hence making decisions that are useful and beneficial for all its stakeholders. The collection and application of such data and related analytics have led to the development of concept of big data. Big data refers to the large-volume data, which is collected on a continuous basis all across the world and usually involve varieties of information that require huge space for storage purposes (Bizer et al. 2012).

Every person in this world, who makes use of technologies, uses big data in some way or the other. When a person is accessing databases from different sources, he/she is using big data, when a person refers to the plethora of information given on Pinterest and other similar sites, it is big data and even when companies analyse the number of customers visiting their websites, and it is again a part of big data (Bizer et al. 2012). Thus, big data has become a part and parcel of the lives of people and is being used by every person. The study thus highlights the significance of using big data for different reasons in a supermarket. The supermarket chosen for the study here is Al- Meera Supermarket in Qatar.

A. Collection of Big Data

Given the size and quantum of information available in the form of big data, it has become extremely important for companies to identify the data that is of relevance to them so that they can focus only on that part of big data and can collect them adequately. The different aspects of big data and the sources from where it can be collected are discussed below.

1) Unstructured data

This refers to the data that exists inside an organisation in an unstructured format. The interesting thing about unstructured big data is that the organisations don’t plan to collect it but it automatically gets collected. For example, the email exchanges taking place to internal or external people in an organisation form a part of big data and can be used for extracting lot of valuable information. Some of the common sources from where unstructured data comes from include emails, documents, social media websites like Twitter, Google Plus etc. and Digital audio and video files (Chen et al.2012).

2) External Data

External data refers to the big data that is generated and collected outside an organisation. There are a number of third party vendors that offer data on different things to their consumers. Data such as the market size, consumer purchase behaviour in the market and even the rankings of websites are all part of external data and can be collected via primary and secondary research activities (Mayor-Schonberger and Cukier, 2013).

B. Industries and Big Data Usage

Considering the immensity and enormity of big data, it can be assumed that all the industries make use of big data, but that is not completely true. Different industries across the world are using big data but the reasons and the extent of usage is different. One of the common usages of big data is to study the market. Industries like telecoms, banking, consumer goods, supermarkets that are highly competitive in nature often use data to understand the latest trends in market so that they can catch up with the changing patterns of consumer behaviour (Zhu et al. 2012).

Industries also make use of big data for highlighting their own market position to others. Use of social media popularity, customer engagement, statistics on the number of consumers etc. can help a company sell itself to other marketers and customers and can be useful to highlight and demonstrate its
popularity. Some of the industries making maximum of big data are found to be energy and resources, life sciences, banking and financial services, utilities, telecommunication, retail and consumer goods (Yoo et al. 2012).

Though other industries and sectors may feel that they do not need big data or that it is too expensive or worthless for them, the fact that it is being used by the entire world and has proven to be helpful cannot be disregarded.

C. Big Data for Supermarkets

Supermarkets are considered to be the superstars of big data because they are among the companies or sectors that extract maximum data and make maximum use of this data. Supermarkets generally have a large number and range of customers with diverse backgrounds. Hence, the data collected by the supermarkets can not only be used by them but can prove to be extremely valuable for other companies as well (Pepe & Pepe, 2012). The ways in which data is collected by supermarkets include the following.

Analysis of the customer database they have provides them with the details of shoppers. For example, supermarkets like Tesco and Sainsbury offer customer loyalty programs, which help in gathering huge volumes of data (Igami, 2011).

The shopping habits and purchases of customers are usually observed and recorded through the IT systems to analyse the goods purchased by them (Reynolds and Walters, 2008).

Supermarkets like Tesco and Sainsbury also offer online purchase options, which help in recording and maintaining data regarding customers (Reynolds and Walters, 2008).

The data thus collected by the supermarkets in these manners provide a lot of information regarding their clients. While the loyalty programs help in obtaining the contact details and personal data and information of the clients, the shopping habits help in analysing the products usually purchased by different groups of people depending on their age, culture, gender, etc (Smith, 2004). The vastness of this data can be judged from the mere fact that Tesco alone has approximately 16 million customers who own the Tesco Clubcard, which means that Tesco itself has access to the database of 16 million people. This data can hence then be used for multiple purposes so that the supermarkets can benefit from them. An example of a supermarket which does not currently use this technology is Al-Meera, Qatar and hence an analysis of how big data can help it in growing itself is discussed.

II. OVERVIEW OF AL-MEERA SUPERMARKET

Al-Meera is one of the biggest supermarket chains in Qatar, which has also associated with the French retailer Casino to open a series of new supermarkets all over Middle East. With over twenty supermarkets in Qatar itself, the supermarket plans to expand and continuously grow in the coming few years and hence needs to focus on the adoption of new technologies and technological systems for better efficiency. The installation of Enterprise Resource Planning (ERP) system has also started and hence the company aims at attracting more customers (Annual Report, 2011).

Despite of all these technological advancements, one aspect where the company still seems to be lacking is big data. As Qatar is a complete gas-based economy, the number of big companies operating in other sectors is very less and hence the concept of big data has not yet seen the growth as it has in other countries. However, the entry of big competitors and multinational corporations has made it not essential to move away from the traditional form of operations and move into a technology based formula. Hence, big data would be a huge, challenging but beneficial adoption for Al-Meera Supermarket (Mohannadi et al. 2013).

A. Benefits Of Big Data For Al-Meera

Increase in the overall profits is one of the biggest benefits of adopting big data for Al-Meera supermarket. As the supermarket can help in understanding the customers’ needs and demands in a better manner with the help of big data, it can then customize its products offerings. For example, analysis of big data would demonstrate the demographics of customers who come in maximum and hence new products suitable to that particular demographics can be added. This would lead to increase in overall sales of the supermarket, thus leading to profits. Also, increase in the number of customers would also lead to higher sales and hence higher profits (Gorton et al. 2011).

As the development of multinational companies is increasing and growing in Qatar, it is obvious that the country is also witnessing the inflow and shift of large number of people from other countries. These expatriates have different needs and demands as compared to the local consumers. Hence, Al-Meera can not only identify their consumption patterns by evaluating big data obtained from the market but can also estimate the number of expatriates and then target them (Wiga and Clarke, 2013).

Customer engagement refers to the process of engaging with customers. Different marketing initiatives are generally used by companies to improve their overall customer engagement and management levels because it then leads to higher number of customers and better customer loyalty. Al-Meera must also try to first understand its customers before devising the marketing strategies so that the effectiveness of these strategies can be increased multi-fold (Kholod et al. 2011).

The supermarket faces myriad of risks related to its finances, marketing and operations. Hence, it is important for the company to try to maximise its efficiency by reducing its risks. These risks can be greatly reduced by the adoption of big data. Al-Meera can not only understand its customers but can also carry out market and competitor analysis with the help of big data. This will reduce the financial risks as financial uncertainties would reduce. Additionally, collecting data on suppliers, processes and technologies used by supermarkets and their benefits can also assist Al-Meera in reducing its overall risks (Kholod et al. 2011). Thus big data can be significantly helpful in reducing the overall risks associated with the business.
While Al-Meera is looking at expanding in different Middle Eastern countries, it is focusing mainly on its current and already possessed expertise. But, there are several opportunities in the market that the company can look at. Al-Meera as one of the biggest supermarkets can thus analyse these opportunities by looking at the data collected from different sources, for eg. if the social media websites demonstrate more discussion on a particular service, the supermarket can try to incorporate that service for higher customer satisfaction (Kholod et al. 2011).

Thus, there are several benefits of using big data for Al-Meera Supermarket. It must however keep the challenges also in mind incorporate the use of big data in an efficient manner.

B. Challenges Of Using Big Data By Al-Meera

One of the biggest challenges that Al-Meera would face in using big data is the lack of skills. Big data needs to be used in a particular manner for making it much more effective and much more usable by the organisation but the supermarket doesn’t currently have any skilled person to carry out this task effectively. Also, since Qatar is not using big data currently, availability of technologies and skills to use big data is very limited and hence it can impose a huge challenge on the supermarket (Mankodiya et al. 2012).

Use of big data can prove to be little expensive for the supermarket and hence from financial perspectives, ROI is a figure that must be kept in mind while selecting and adopting such changes. Considering the investment, it is possible that the ROI that is obtained in the short term is not very high and hence that imposes another challenge on Al-Meera supermarket. In fact, there is not even a definite way in which ROI can be accurately measured. Thus, the supermarket needs to find a way around this deficiency as well (Hemerley, 2013).

The societal factors and culture that exist in Qatar are extremely different from those observed in the western countries. People in Qatar are found to be extremely religious and also more conservative as compared to people in the western countries. Thus, these social and cultural factors can impose a challenge on Al-Meera supermarket while collecting the data (Wigan and Clarke, 2013).

Thus, there are numerous challenges that Al-Meera supermarket needs to witness and tackle while incorporating and using big data for its organisation. Overcoming these challenges may be taxing but can prove beneficial in the long run. Strategies that can be adopted by the supermarket for making maximum use of big data are discussed below.

III. STRATEGY RECOMMENDATIONS FOR AL-MEERA

There are several strategies that can be adopted by the supermarket in order to make better use of big data and its analysis. The first measure that should be adopted is the recalculation of its risks related to the use of big data. In order to effectively employ and use big data by Al-Meera supermarket, it must recalculate its risks in terms of minutes. This will allow the supermarket in taking calculated risks and the risks and benefits of big data can be weighed against each other. This would help Al-Meera in effectively understanding and managing its risks (Morland et al. 2006).

There is a huge scope of big data for customer management and customer collaboration. Every supermarket has two types of customers, one time customers and the customers who make repeat purchase and are loyal towards the supermarket. Hence, Al-Meera must use big data and try to identify the customers who are most important. The customers who make repeat purchases and are loyal towards the supermarket would be the most important from the company’s point of view. Hence, the first focus of the supermarket must be the identification of important customers (Kouris et al. 2005).

With the increasing competition, it is becoming more and more difficult for companies to increase their brand awareness and loyalty. Increasing the number of customers is an important aspect of an organisation and hence one way to achieve the same is by providing them with customized solutions for meeting their personal needs. Al-Meera can also make use of the big data for sending customised and personalised recommendations to its customers. Based on the data that is obtained regarding the buying behaviour and preferences of the consumers, the supermarket can form personalized recommendations for them, which would not only increase sales but can lead to higher trust, brand loyalty and engagement in customers (Goldberg et al. 1999).

Test Campaign: In order to determine and establish the effectiveness of big data, Al-Meera can also make use of test campaigns. It can devise a marketing strategy based on big data analytics and can then calculate the ROI and profits and benefits obtained from these marketing strategies. This would help in estimating the exact benefit of using big data (Goldberg et al. 1999). Test campaigns avoid the huge expenditures that are required for actual implementation and are best when the surety of a plan being effective is not present.

These recommendations are some basic ways by which Al-Meera supermarket can enhance its overall outcomes and use of big data in the most effective way possible.

IV. CONCLUSION

Following the likes of western country supermarkets like Tesco, Sainsbury etc., it is important for Al-Meera Supermarket to also use big data for increasing its profits. The analysis presented above clearly highlights and proves the fact that big data helps in growing the company even bigger an hence looking at the expansion strategies and plans of Al-Meera, using big data can be a smart and effective move.

V. REFERENCES


Big Data For Insurance

Ali Al-suwaiedi
School of Computing and Mathematics
University of Derby
Derby, United Kingdom
100310963@unimail.derby.ac.uk

Abstract— The ever-increasing flow of information in organizations has posed challenges to business for years, and this has led to the development of the phenomena of big data. Its development has helped in revolutionizing management of information including dissemination in diverse areas of operations within a setting. The term big data has used has diverse elements under which scholars for its description that include velocity, volume, veracity, and the variety. For years, companies such as insurance firms have correlated data sets using the five elements. This has helped limit confusion attributable to the collection, storage, and dissemination of complex data. The correlation is necessary due to the difficulty in managing huge volumes of data sets. This paper gives credible information pertaining to big data that refers to huge volumes of multifaceted data sets stored in a database purposely to aid the execution of diverse activities depending on the nature of the operations of the organization. It covers definition of big data, its major V’s, historical aspect, its opportunities in improving operations in the insurance sector, benefits, and support systems including recommendations. The information is availed to enlighten major stakeholders in the information to understand the imperativeness of big data, and how it boost data security.

Index Terms—Big Data, Insurance, Governance, Privacy, Small and medium enterprises.

I. INTRODUCTION

Today the term big data is well known, but many may think that it will only be existed and benefit large organization, the fact is that big data may exist in SME’s and they may gain many benefits from it. The main challenges that face the organizations are sources of big data, analyzing big data, and the privacy issues associated with big data (Davenport and Dyché, 2013).

II. WHAT IS BIG DATA

Big data refer to huge volumes of complex data sets stored in a database to help the execution of diverse activities that is needed based on the nature of the operations of the organization. The major type of information stored in the data sets includes structured data and unstructured data on companies, people’s profile, and research findings. In large organizations, the correlation of the data sets is essential in minimizing the complication attribute to the difficulties in collection, storage, and dissemination of huge volumes of information from different sources. (Dumhill, 2012) the correlations in sets minimize the confusion in management of data. In particular, the extent of the size of the data makes it difficult for businesses to utilize traditional management tools in processing the data.

A. The five V’s of Big Data

Big data was defined with initially three V’s and it was changed up to 12 V’s, here only five are relevant, Volume is one. It describes the enormous amounts of data sets that is produced per second. The measurement for the quantities of data is in bytes. The other two V’s of the five V’s are velocity and variety. The last two are veracity and value. For the velocity, it is the pace of creation of new data defines big data. From the definition of volume, data is rapidly expanding and as such, organizations will face difficulties in managing volumes of data in years to come. This will increase the speed of production of data. As a V of data, velocity influences the capacity of storage of data in databases. Variety is another V that facilitates systematic analysis of individual’s capacity to generate data. Variety describes the uniqueness of data in terms of the data structure. As noted, huge volumes of data transmitted in different parts of the globe are unstructured (Alex, 2013). For veracity, it is the accuracy of data and is the data is correct or no. Value that refers to the usefulness and effectiveness of the data forms that last V. The estimation of value is dependent on the significance of the data (Dodge, 2013).

III. HISTORY OF BIG DATA

The exact origin of the use of the term “big data” is difficult to establish, however, it is evident that the growth of internet technology has contributed significantly into its existence. Notably, the year 2012 was when the growth of big data became evident as people were seeking solution to the limitations in data expansion and usage. Organizations and individuals at the time were searching for effective tools for capturing, processing and dissemination of huge volumes of data leading to the development of the phenomena of relating data sets. The objective was to deal with the additional increase in information flow. It would help ease the flow and management of information in organizations and companies. In 2012, it was revealed that big data was established under three different categories such as technology, analysis and mythology. Technology that aimed at the use of computation in organizations to gather analyzes link information and compares huge data sets (Rijmenam, 2013).

The next category was analysis, which involves scrutiny of large data collection to determine a certain pattern in order to
make social, technical, economic, and legal claims. The last category is mythology that deals with widespread certainty, which involves large data sets that provide higher form of knowledge and intelligence that can be used to come to a conclusion on information that was previously impossible explain but would rather be explained with facts, objectivity, and accuracy. The below diagram presents the history of big data development; it was started in the year 1989 with FICO company launched. It gradually moved to the social era, which was between 1994 and 20013, which introduced various ideas such as Amazon, Expedia, Google, Moneyball, Twitter, iPhone, Facebook, Adobe, EMC, and FuelBand among other respectively(Liewehr, 2013).

![History of Big Data](image)

**Fig. 1. History of Big Data**

**IV. OPPORTUNITIES FOR BIG DATA TO IMPROVE INSURANCE**

Insurance has the capacity to reduce their expenses in using big data in managing data for its clients. All the activities in an insurance company depend greatly on data management, which can at times be a lot of work to accomplish manually. Through customization of services, small insurance companies have the opportunity to use big data of large companies to design packages that suit the needs of different clients of insurance companies. It has been established by experts that the use of big data will have positive effects on businesses especially on insurance companies. The technology will be used in insurance companies to make sense of the internal and market information and records that are obtained in the operations of the insurance companies. This will help the companies to achieve competitive advantage over other companies. This is because the technology will improve a company’s overall performance by facilitating deeper relationships with customers, greater pricing accuracy and more efficient and effective loss prevention (Kenealy, 2014).

This is an opportunity for insurance companies to improve on their performance and gain a greater market share. Small companies can also alter data sets to enable the management to assess the situation in the market place prior to designing premiums that reduce wastes. For instance, insurers in small firms will liaise with actuaries from big firm when pricing their policies to facilitate customization of premiums. The application of big Data will give insurance companies a platform to manage their data, which will enable them work efficiently and effectively. The use of this technology will present insurance companies to have advanced analytics, which will help them reduce the gap between large and small corporations. It is due to lack of analytical capabilities among smaller insurance companies. This will ensure that all the companies perform better in the market (IBM, 2014).

**V. CHALLENGES TO USE BIG DATA FOR INSURANCE**

Every opportunity that presents itself is always accompanied by challenges. This is the case with big data technology; in its great efficiency and effectiveness, there are challenges that still affect this technology. It is difficult to validate big data since they are stored in subsets obtained from correlated sets of data. The ownership of big data is also unknown since it come from various of sources that sometime it is difficult to know the source , and this makes it difficult for organizations to process big data in real time, in situations where the sources of the big data is unverifiable and unknown. Subsequently, managing big data requires huge investments, and this poses a challenge for small organizations that lack the capacity to extract important data from huge volumes of data. Breach in information is the greatest challenge that presents in the technology. The information that has been stored in the system might be breached by unauthorized persons, which might lead to the loose or access of personal information such as pins, names, credit card information among other details. Any breach of information has great impact on a company as they might lose clients or even sued by their clients, which will at the long run affect the performance of the insurance companies (Robinson, 2014).

**VI. BENEFITS FOR INSURANCE TO USE BIG DATA**

Big data advances analysis of diverse issues that helps in the identification of the major causes of an operational complication in organizations at early stages thus preventing the possibility of huge damages in the future. The real time assessment of data encourages customization of products to suit the needs of different customers. This saves insurers costs incurred in researching the marketplace during the process of calculations for risks. Consequently, the real-time analysis supports in identification of new opportunities in the marketplace. For instance, small companies can seek current and potential customer information using mobile phones, social network, and other sources thereafter use the data in pricing packages and minimizing the potential risk, and to make suggestions that can fit the customers need(Saama, 2014).

**VII. OPERATIONS NEEDED TO GET AND ANALYZE BIG DATA FROM INSURANCE**

The use of effective technologies in the collection of data is a critical step in ensuring the retrieval of useful data. Proper timing is also necessary to prevent data from spinning. The speed of collection has to secure the back-up systems. The confidentiality of customer’s information is also mandatory. Consideration of the above factors aids in leveraging data based on key elements of big data that include variety, velocity,
and volume among the other Vs. Operations in data analysis should ensure contextualization of big data from diverse perspective to facilitate integrative sampling in order to enable small organizations use data of large organizations in making deductions (Hurwitz, J. 2014).

VIII. GOVERNANCE OF BIG DATA

Data governance refers to a strategy that makes sure that vital data assets of a given company are managed in an organization or company. This process ensures that company data is trusted and made accountable, which provides quality for integrity information. Data governance is about applying technology of different capabilities to help create smooth operations within an organization (Drenik, 2014).

Governance of big data is available and should be done in respect to the organization that has implemented it in its daily operations. Big data offers many companies the opportunity to know or have the statistics of the clients and operations. It is upon this that companies have employed data governance techniques in order to protect their data. This helps in ensuring that all the information that exist within an organization are governed (Borodzicz, 2009, p.53). A company that has already applied data governance is IBM, which wants to ensure the integrity and governance of information in the company (Elliott, 2014).

IX. PRIVACY AND BIG DATA

Privacy is a process through which an organization’s information is secured against unauthorized access. This is done by considering the most likely ways through which the information can be tempered with or leaked and breached. Privacy and data governance is vital in ensuring protection of information that a company owns. This is because of the different activities that take place such as online data transfer, multiple transactions among others, which might lead to risk and this is why privacy is ensured. IBM takes privacy and data governance as one of their major concern in their business environment (Drenik, 2014).

X. SOCIAL NETWORK AND BIG DATA

There is a big relation between social network and big data; there are many internet users in the world today that take part spend more than half their time on social networks. It is upon this data the online social media companies use it to deliver great content to individuals according to their preference. Big data is used in the social media to have the best of content and information that they get or give to their consumers (Smith, 2014).

XI. HOW INSURANCE COMPANIES CAN OBTAIN BIG DATA

The insurance companies obtain big data through the records of the insured parties and the high number of premiums registered with such companies. The high number of parties insured and covered by the insurance companies leads to massive data on the systems of such insurance companies. The different classes and types of risks covered by the insurance companies lead to very large amounts of data kept by the insurance companies. The records of the premiums paid by the insurance company also adds up to large volumes of data handled by the insurance companies. The data of the claims by the insured parties also contributes greatly to the large amounts of data kept by the insurance companies that creates big data (Adams, 2013).

XII. DOES INSURANCE COMPANIES HAVE BIG DATA?

The insurance companies have large volumes of data relating to the premiums paid and the claims lodged by the insured parties. The data relating to risks covered and the investigations carried out to prove the existence of such claims leads to very large volumes of data handled by the insurance companies (BOULTON, 2014).

XIII. HOW CAN INSURANCE COMPANIES USE BIG DATA?

The insurance companies may use such big data in various ways in order to facilitate their operations. The insurance companies use the big data to analyze the aspects relating to the levels of risks in order to determine the amounts of premiums to be paid by their clients. The data also help the insurance companies to determine amounts of claims to be paid so as to avoid cases of losses by the company. The insurance companies also use big data in order to simulate and predict the chances and probabilities of the risks that are likely to occur. Big data helps such companies to forecast uncertainties in the futures as well as formulating appropriate strategies to handle such risks (Barnes, 2014).

XIV. RECOMMENDATION

It is advisable for insurance companies to invest in big data in planning their administrative duties to limit wastage of resources that result from the use of the inappropriate strategy in meeting the needs of customers. Big data is essential because it unearths opportunities besides encouraging collaborative work in assessing market dynamics. Users of big data should be cautious in their actions to avoid challenges such as security risks and bias in the collection of data. Insurers should also
scrutinize the data in details to minimize the possibility of incurring losses from using wrong data in making decisions.

XV. CONCLUSION

The use of big data is a new phenomenon in the business world, hence the increase in its usage despite some of the data creating problems for users. One of the benefits of big data use that has increased its usage is its efficiency in facilitating analysis. Effective harnessing of big data also reduces wastages, saves on time and enhances effective decision-making. On the contrary, users of big data face challenge in designing effective database management systems for huge volumes of data.

XVI. REFERENCE


Implementation And Data Latency Issues Regarding Big Data

Mathew Bateman
School of Business, Computing and Law
University of Derby
Derby, Derbyshire
M.Bateman4@unimail.derby.ac.uk

Abstract: This report will look at how you could efficiently utilize big data analytics. The variety, volume and velocity of the data has been considered whilst checking the accuracy of the data that is actual provided via the user interface from the end user. Following this the report proceeds to look at how latency can massively affect the usefulness of the data that is gathered due to the time frames making data irrelevant.

Keywords: Big data, analytics, implemented, easily, efficiency, revenue, stream

I. INTRODUCTION

Throughout this report I shall be investigating how efficiently big data analytics can be used in small to medium sized business thus in turn having an effect on the company’s revenue stream.

Big data analytics is the common definition for the analysis of datasets that are of a large scale that become near on impossible to humanly decipher from techniques such as conventional data management practices to traditional processing of the data. One major problem that current data scientists face is the scale of the collected data, however, this larger dataset is usually the one that reaps the most rewards due to it allowing the detection of relationships between business trends and business practices.

Companies currently store for the large part their information in a relation database management system the draw on statistical programs and visualization packages to process the majority of the work and make the data available for humans to understand and action (Hsieh et al, 2013). However, big data frequently relies on the power on parallel software that’s backed up by thousands of computers. Size and how to process the data is becoming ever more apparent to business enterprises with for Walmart handling more than one million customer transactions per hour (Washington Post, 2014) which is equivalent to 167 times the information in the US library of congress hour (Washington Post, 2014). Making this information useable, understandable and available in a timely manner can present a huge cost to the businesses.

I shall begin with looking at how user interactions with retailers have changed with 56% of American adults now owning a smart phone (Smith, 2013), and how this will massively increase the reach of a company’s target audience. Then I shall look into how an SME can use its people to make the most efficient use of this data whilst not having a negative effort on wage costs for the company. Overall I believe that SME’s will have a positive effect from implementing the use of big data analytics in their companies to maximize their revenue streams in a positive way with examples of poor data management often costing the company 20-35% of their revenue streams.

One aspect that I shall be investigating is how employers can implement the use of big data analytics into their company. Firstly though the cost aspect to the business in relation to their potential rewards which may often come across as very limited. Secondly I shall go on to look at how they can employee people to efficiently manage and analyze the data, through the discussion of just one data scientist or to a team of data scientists with each being responsible for an aspect of the analysis process for themselves due to their being many aspects of big data analytics that if a business wishes to utilize all aspects of the information they gain from the data cannot be fully interpreted correctly by one single user due to the complex nature of the data.

Another aspect I shall investigate in my report is how can the data be kept applicable to the business in hand, due to many companies now being to present restrictions on the sorts of data that companies collect on their users due to increasingly be wary of their privacy and how deep and accurate big data can actually be in representing a customer’s life style and habits and if this is even appropriate for a company to be collecting.

Following on from the point above I will investigate how big data is likely to change in the mere foreseeable future, which in this subject area has a massively predicted 40% global growth in the creation and storage of big data by companies with only a 5% increase in IT spending globally may see an impact upon the successful implementation of big data due to companies not being willing to utilize such readily available information.

Finally I shall be looking at how easily the technology is available to the businesses looking to implement big data analytics and the ongoing support and progression with these software packages

II. USER INTERACTION

With modern adoptions in technology and user culture many people have recently been upgrading their mobile phone to smart phones that have the power to access the World Wide Web and access the many web portals of companies.
According to a study by the pew Research Center’s Internet & American Life Project approximately 56% of all Americans have a smart phone (Smith, 2013) this new advent of smart phones provide many retailers with a direct method of monitoring their users whilst in store or their online buying habits from the company.

Not only are users being able to access retail environments from their smart phones many use them as a means of social networking with the major players being Facebook and Twitter, both have commissioned reports into their inner process and these reports have indicated that Facebook has thirty billion pieces of information shared on their platform monthly (Mckinsey.com, 2014) With Twitter indicating that it sees roughly 175 million tweets from more than 465 million account in 2012 (Dawson and Ziv, 2014).

With the examples discussed above it is with no surprise that the world’s largest retailer Walmart generating approximately 2.5 petabytes of data being generated through its point of sales across their company every hour (Big data meets big data analytics, 2014, p. 3). These examples really demonstrate that data is being generated on a massively large scale that could be readily available and accessible to the business with the right big data analytics in place. There is evidence as well to suggest that even though consumers actively search online for the best deals they can get often revert to purchasing the goods they sought in store. (Washington Post, 2014) This shows that even when big data implantations don’t actually distract direct business away from the bricks and mortar stores to the online environment.

This shows that getting customers to interact on a digital level isn’t actually the difficult part for a business to implement big data analytics isn’t the difficult part. However, as I will go on to discuss to getting the data into an understandable and actionable form for the business.

III. IMPLEMENTATION COST TO THE BUSINESS

To begin with I shall discuss how the implementation of big data directly impacts a company financially. Obviously to implement big data analytics there will be a direct cost to the business. However, this may actually not be as significant as many business expect. One of the major players in big data analytics software being Apache Hadoop (Sys-con.com, 2014) which has reports of 94% of their users are currently performing analysis on larger volumes and greater detail than previously thought possible (Sys-con.com, 2014). Its compression algorithms allow the approximate 82% of the data that was previously not storable now accessible to the system and the data scientist (Sys-con.com, 2014) Hadoop as discussed then provides very real advantages to a business of any scale that decides to implement big data analytics at a massively reduced cost because the software is open source it doesn’t actually represent a cost to the business implementing it.

Another major advantage to the open source nature of Hadoop is that it allows collaboration on problems that the software does not currently provide to be worked on by the large user base of the open source network (Sys-con.com, 2014). However, this does mean that specific and timely support of the software may be very difficult to secure and as such an in house specialist may be required to provide real time support to the company so they don’t lose out.

IV. DATA LATENCY

Another issue to consider a problem with big data is that the data may not be collected quickly enough (School, 2014) This may provide organizations to be at a major disadvantage and not allow them to react in a timely and efficient manner to trends that may be predicted via the use of social media big data.

To further expand on the point that was highlighted above and the latency for Big Data being an extremely important factor on the reliability of the validity of the data (Dawson & Ziv, 2014). How to address this is becoming a more than critical aspect of the development of big data with some research being conducted into how the system architecture could be used to improve on the goal of low latency between the data itself and the machines that process it. (Jun & Liu et al., 2014)

The paper mentioned prior looks at how an architecture can be used to achieve a lot of data being passed through the system by using a significant amount of flash memory managed by flash controllers. (Jun & Liu et al., 2014) Due to example of the aforementioned using flash memory and the ever growing mass market for flash memory, the industry is greatly increasing the development of the flash memory technology ranging from USB flash drives to solid state drives that are capable of powering servers (Hsiew & Chang et al., 2013). This technology has shown greater enhancing the performance of systems ran on this technology allowing them to process data far more efficiently and also in a very fast manner (Hsiew & Chang et al., 2013).

To a company that is employing this technology in its servers processing the big data, may run find the technology to be very problematic for many reasons. One of the main reasons has been the reliability of the hardware with after approximately 10,000 read/writes the devices performance starts to significantly suffer with the predictability of the underperforming blocks of data being extremely hard to predict due to it being specific blocks and no specifically the whole hardware (Hsiew & Chang et al., 2013).

Furthermore from the implementation of the flash memory storage allows for a further method to decrease the latency between big data and the processing of it is a computational module being located in the flash controller located between the storage and the host, this allows the system designers to allow for hardware acceleration that processes the data without any additional latency for the data (Jun & Liu et al., 2014).

This whole aspect of flash memory help decrease the latency of data processing because the whole data is often too large to be cached in the main memory of any host at a reasonable cost to the end user, and to be able to fragment this data into a fast read technology then this drastically improves the performance of the big data to the end user, however provides a system that at any point in time may fail due to the
nature of flash memory and due to the infancy of flash memory a very high cost to the end user (Jun & Liu et al., 2014).

This shift in progress from the weakest link being the storage device due to the improvements of flash memory moving the focus across to the network infrastructure itself. As a result of this system designers now have to focus their attention onto utilizing and optimizing all aspects of the systems from software to storage and network infrastructure underpinning the system (Jun & Liu et al., 2014).

This leads directly into shifting the data from batch processing into streaming the data as and when it happens, thus rewriting completely how we currently process big data. This has lead to a new distributed system that has design aspects specifically targeted at low-latency continuous process of the big data on standard of the shelf computers at a relatively low cost to the business user (Qian et al, 2013). This approach has drastically changed the design away for the popular MapReduce style of batch data processing to improve the latency response of the system. Traditionally the data is processed in batches when a limit is reached. The new suggested style of processing allows the continuous processing of data as and when the system receives it, this new method echoes how big data changes more precisely whilst bridging the gap towards storage on a database too(Qian et al, 2013). Due to the nature of processing the data when the system receives it, the system must be capable of coping with temporary changes in load distribution as well as failure and recovery (Qian et al, 2013). This new style of processing design allows the analysis of Twitter data at a peak rate close to 10,000 tweets per second with approximately two second delay (Qian et al, 2013) the sort of data size and response time required closely represents the big data needs of current business and this new style of implementation may allow them to get the most reward from big data in a timely fashion in an affordable manner to their profitability.

One aspect to greatly consider whilst looking to implement low latency in big data systems in fact is the very architecture on which they run, providing a very impact on the latency of the system in practice. One major example of this is Twitters implementation as discussed in the report “Fast Data in the Era of Big Data: Twitter’s Real-Time Related Query Suggestion Architecture” (Mishne et al, 2013) The reasoning behind this being so suitable to that of low latency is that Twitter process their Tweets on a real time basis that closely resembles information and data is received and processed in a real world environment and as such has very transferable aspects to that of a business (Mishne et al, 2013). Twitter’s first implementation of processing tweets was via that of an Hadoop-based analytics stack (Mishne et al, 2013) This proved troublesome on allowing them to handle the big data in as fast a way as possible to allow for precise and relevant manner (Mishne et al, 2013). To overcome the cumbersome style of Hadoop processing meant that it provided the company with a high latency that although provided great experimentation however didn’t deal with the current definition of big data (Mishne et al, 2013).

As discussed many times above the traditional platform of Hadoop based processing of big data doesn’t provide a very real and appropriate representation of the data (Zaharia et al., 2013). The current Hadoop based platforms however handle fault tolerance and recovery extremely well due to its ability to replicate (Xue et al., 2012). However as highlighted by many of the articles above a real time processing system doesn’t provide a very efficient way to deal with fault recovery in a none expensive manner like the alternative Hadoop style does (Zaharia et al., 2013). The report mentioned prior goes on to suggest a method that could be used to adequately deal with fault tolerance and replication via the use of streams that can be used to overcome these challenges (Zaharia et al., 2013), these allow the process to be ran in parallel allowing one to be the active and used data whilst the other is stored and as such allows traditional replication, backup schemes and allows improved efficiency (Zaharia et al., 2013). The process goes on to describe how it creates a hybrid between the two to gain the most benefits from both real time processing and the map reduce processes of gaining the relevant information from the raw data (Zaharia et al., 2013).

The major increase in fast data through the use of social networks and world applications ranging from the stock market to sensor data streams. MapReduce has emerged as a very capable process of dealing with the data where a developer writes a map function and reduce function and the system allocates the resources where appropriate (Lam et al., 2012) These are then separated and referenced according to their data due to the continuous nature of the data slates are the raw data in a stream at any given point (Lam et al., 2012). This proves problematic in the nature of the data because these reference the data at a given point in its life span in the system compared to stream computations where the raw data is frequently changing and needing to be updated thus changing the learning outcome of the data also with MapReduce there is no such thing as a start and an end to a stream it is by definition a continuous flow of data. Due to the map reduce model requiring a key in the reduce function (Lam et al., 2012) to appropriately action it’s procedure this provides a challenging aspect on the storage of this information in a manner that can be accessed and updated fast and efficiently (Lam et al., 2012) Another major flaw with the implementation of Map Reduce procedures on stream computations is that you may restart the computation of a map reduce procedure from scratch (Lam et al., 2012) due to the nature of a continuous stream of data this will prove a very challenging point for the system to overcome and process again in an appropriate and timely manner (Lam et al., 2012). All of these problem suggest that the use of Map reduce in a continuous data flow isn’t going to gain the most efficient and relevant information from the data.

V. Conclusion

To conclude, many aspects of big data were discussed during this report ranging from the implementation cost to user interaction between a system and big data. Finally I went on to discuss the factor of latency in big data and how in real world
applications such as the stock market the mere milliseconds of processing required would carry major financial implications.

Ultimately big data will have to be implemented in some form or another to any company in the foreseeable future be it for processing their transaction history or predicting their utility use it will help define almost all decisions made by businesses big or small.

A big part of the research was that into latency and how it is present in the current implementations of big data processing and how different approaches are being used to try and massively reduce this to mere nano seconds of processing before the raw data becomes usable in a real world situation such as making trade decisions in the stock market. Many different ways are being used to reduce this such as machine architecture to network infrastructure and allowing the data to be streamed in a semi real time fashion, replicating how it happens in the real world. These all present challenges to the technology we have available at the moment and how we can stretch this to provide a transferable and relatively low cost to the business which would encourage them to include big data to their business in a more timely manner. All the method highlighted in the research such as architecture redesign and even procedure redesign due to the nature of the data changing show that this is a changing field where the technology is changing to match the requirements of the data and as at present to reflect this on the original question and how efficiently it may be to implement to a business, the quick changing nature may prove to be challenge or even off putting for a business to implement it in the near future.

VI. REFERENCES


Parallelization Considerations For Big Data Analytics

William Briggs
Dept. of Computing and Mathematics
University of Derby
Derby, United Kingdom
w.briggs1@unimail.derby.ac.uk

Abstract—Data these days comes in large quantities, as implied by the term ‘big data’. Some data is so large it would be almost impossible for a single machine to compute and process the potential millions of requests made to it each day. The solution to this is to parallelize the processing across multiple machines in order to improve the performance and responsiveness of the entire system. This proposes numerous considerations for people responsible for building and managing big-data analytics software.

Keywords – Big Data, Analytics, Analytical Software

I. INTRODUCTION

This article is going to investigate into the various API’s available for processing large quantities of data and the suitable ability of parallelizing the computation. It will also discuss what hardware considerations need to be made when implementing and maintaining big data analytics systems.

Defining what constitutes as big data is somewhat debatable. Most people define big data by the size of the content being processed; such as terabytes – and now even petabytes (Russom, 2011). Some other organizations define big data in terms of time (Russom, 2011). Of course, defining big-data by size could introduce a level of ambiguity in a time where storage device volumes are ever increasing.

A much more suitable model that exists for defining big data is known as the three V’s. These stand for volume, velocity and variety. The reason for this definition is to defer the myth that big data is only about the size (Russom, 2011). There is evidence for this when writing an application that processes big data, if the variety of the data is unstructured then we need to cater for this.

These are of course many methods to structure data with dependence on the attributed filter. One algorithm which is based upon structuring computation while enforcing parallelism is merge sort. It works by splitting the data into equal segments and placing each segment in its own array. Each array is then sorted in parallel, before being merged back into one at the end.

Algorithms like merge sort are fairly outdated and there are often more efficient way of distributing workload that will produce quicker and more reliable results. Furthermore, implementing merge-sort; or a similar algorithm could be a tedious task. However, we still need to be able to distribute our computation in order to process large quantities of data.

Parallelism is important in this day and age due to the fact we have reached a point where CPU manufacturers strive to incorporate more cores into their products over increased speed (Hill, 2005). The reason for this is due to the physical constraint applied from cooling and power consumption of a high-clock speed running processor (Sun and Chen, 2009). A law known as Amdahl’s Law governs the speed increase that can be achieved from parallel processing; it is seen to be a fairly simple and elegant law (Krishnaprasad, n.d) that can quickly be used to determine how much performance increase can be acquired from task parallelism. The graph below shows how much speed up can be achieved by making use of Amdah’s law.

II. DISTRIBUTED SOFTWARE DIFFICULTIES

Most companies consider timely analytics over big data to be a key ingredient for success in business (Herodotou et al, n.d). A single machine is unlikely to hold hundreds of terabytes of data on its hard drive locally; let alone petabytes. Furthermore, even if a single hard drive could hold petabytes of data the CPU and memory would never be able to analyze the data at an efficient enough speed for a company to get results in a timely manner. This suggests parallel/distributed storage and processing of large quantities of data in a timely manner is a consideration companies must make when wishing to engage in big data analytics.

Let us now consider the difficulties in building and debugging distributed software.

Writing software that is able to run in parallel is known as concurrent programming. Although it is generally conceded debatable; concurrent programming grew out of problems associated with operating systems (Ben-Ari, 1982).

For a processor to give the impression it is performing multiple actions at once it will employ a procedure known as context switching, this allows for the underlying operating system to shift between tasks at an exceptionally fast rate,
allocating resources to each process as it switches. Even though a single-core processor can never achieve true parallelism, the rate at which the operating system interleaves operations means it is fair to assume tasks are being performed simultaneously (Ben-Ari, 1982).

Pretending processes are being executed in parallel means it is easier to write code that truly exploits a multi-core CPU. This is due to the fact the mathematical treatment is simplified if we impose an order of the instruction that are compatible with shared execution (Ben-Ari, 1990), in other words code that is meant to exploit a multi-core CPU is often treated the same as code that makes use of a single-core CPU with task switching.

There are numerous problems that arise from concurrent programming; as hinted by the numerous texts covering the topic. Some examples of problems introduced with concurrent programming include: deadlock – where the system stop responding and data races where the action of the system is dependent on which thread finishes its processing first.

III. SOLUTIONS

While distributed programming is still considered a difficult topic to comply with, (Alvaro et al, 2010) there are various API’s and interfaces provided that help to write distributed applications with less concern over thread coordination. These tools raise the level of abstraction for programmers and allow for less concern over thread coordination and more emphasis on applying functional and logical expressions to collections of data (Alvaro et al, 2010).

One form of API which has become one of the most widely used in industry (Dean and Ghemawat) is MapReduce. MapReduce is an easy to use, scalable framework that allows for data processing to be computed in parallel. I believe part of the reason for MapReduce’s success is down to the design of modern processors where multi-core architecture has become a norm.

Before Map Reduce it was common to see hundreds of special-purpose computations (Dean and Ghemawat, 2004) that processed large amounts of data. The data that was being processed started to become large and required for the processing to be distributed across multiple machines (Dean and Ghemawat, 2004). Google’s first technical response to the challenges of analyzing large amounts of web data was known as the Google File System (GFS) and is known for starting the big data revolution in the systems world (Borkar et al, 2012). This provided developers a familiar OS-level abstraction, but big data often still contained within a very large file, which was often too big for a single machine. As a result of these problems MapReduce was developed which allows for computation parallelization and provides the characteristics important to a distributed system (Ghemawat, 2004). This means developers can write software that processes large amounts of data, where the contents of the file are distributed across multiple systems.

Maps reduce works by defining two main functions – map and reduce. Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs and for the reduce function merges all intermediate values associated with the same intermediate key (Dean, 2004).

One of the biggest advantages of writing applications in this style is that they are automatically parallelized and the computation can be distributed across multiple machines (Dean, 2004). This characteristic means it is significantly easier to write an application that processes large quantities of data, and, because computation can be parallelized, the results can be achieved in a timely manner; regardless of the volume. This ability to scale the processing with dependence on the job size is another key attribute of Map Reduce and means companies can just add more servers if the processing jobs become larger.

Another type of API that was significantly less successful than Map Reduce; but shared similar ambitions was a project call Dryad by Microsoft. Dryad is a general purpose distributed execution engine for coarse-grain data-parallel applications (Isard et al, 2007) and was developed in response to the Map Reduce (Borkar et al, 2012).

The Dryad project had very precise and decisive aims: Make it easier for developers to write efficient parallel and distributed applications. The motivation for the service was the emergence of large-scale internet services that depend on masses of general-purpose servers (Isard et al, 2007).

Ultimately the Dryad project was not a success and in 2011 Microsoft discontinued the work on Dryad and decided to place their emphasis on the Hadoop framework (Foley, 2011).

These types of big data analytic frameworks, which are normally available open-source, are now used by many large companies. Facebook uses Hadoop to analyze their data, and create Hive, a data warehouse system. eBay uses Hadoop to optimize their search results and Twitter uses Hadoop for log file analysis and other generated data (Saecker and Markl, 2013).

A. Framework Limitations

The use of frameworks like Map Reduce might seem like a sufficient and elegant way of distributing the processing of files containing big data. But there are some other considerations to be made.

B. Map Reduce limitations:

- No global security mechanisms provided for protecting sensitive data (Hamlen, et al, 2010).
- Oriented layout – The row oriented layout of map reduce has caused some problems due the oriented
design (Dittrich and Quiane-Ruiz, 2012). This is because additional queries have to be sent through the network in order to merge different attribute values into a row, this is known as tuple re-construction (Dittrich and Quiane-Ruiz, 2012). Sending additional content through the network can increase the costs.

- Assumption that the job can be parallelized, but, some tasks, such as data processing, this might not be achievable.

C. The security problem highlighted above can be instigated by the facts:

- ‘Any business running a Hadoop cluster gives all programmers and users the same level of trust to all the data that goes to that cluster.’
- ‘Any job running on a Hadoop cluster can access any data on that cluster.’
- ‘Any user with limited access to the jobs they can run, can potentially run that job on any data set on the cluster.’
- ‘Malicious users could modify other user’s data.’

D. Parallel DBMS vs. Specialist Parallel/Distributed Programming Models and other Considerations

Interestingly, an alternative to Map-Reduce-like tools has existed for over 20 years (Pavlo, 2009). Parallel database management systems are high performance and robust systems that often provide a high-level programming environment suitable for writing parallelizable big data analytical applications (StoneBraker, 2009). Example products include: Netezza, Paraccel and Datauapia. There are numerous considerations when deciding upon which system is best suited. Parallel databases; like regular databases, use the SQL language for query execution. SQL is considered a more understandable and easier to understand language than Java (Mchome, 2011).

One of the most important factors to consider is the performance of each system. Parallel databases are shown to be significantly faster than Map Reduce; especially for the join task (Mchome, 2011). This performance advantage is credited to the number of technologies developed over the last 25 years, including B-tree indices to speed the execution of the selection operations, novel storage mechanisms, the ability to operate directly on compressed data and sophisticated parallel algorithms for querying large amounts of relational data (Pavlo, 2009).

Another consideration is the usability of the system. Map Reduce has the advantage when it comes to setting up the system (Pavlo, 2009), while setting up a parallel database management system is considered significantly more difficult (Mchome, 2011). However, SQL queries are still considered much easier to write than Java Mapper and Reducer classes (Mchome, 2011).

In terms of reliability Map-Reduce has the edge. This is due to its ability to recover from hardware faults with little overhead. However, this benefit comes at some cost due to the cost of materializing the intermediate files between the map and reduces phases (Pavlo, 2009).

Other technologies that we have also seen the emergence of are specialist programming languages designed processing data. One such example is the R programming language which over the last decade has become the single most used tool for computational statistics, visualization and data science (Smith, n.d.). However, with a focus on more parallelized systems, R might start to lose its popularity due to having limited support for writing parallelized software (Trelles, 2011).

IV. HARDWARE CONSIDERATIONS

The choice of software and the type of API’s used are important when it comes to parallelizing the computation of big data, but the underlying hardware can also have a significantly impact on performance too.

As the volume of big data is raising (Saeccker and Markl, 2013) businesses do not want to invest in hardware that will soon be outdated and not be able to perform the tasks they require.

It currently takes around 9 hours for a 1000 node cluster of machines to each process 500GBs of data, totaling up to 500TB and is estimated to cost $3,000 (Trelles, 2011). A less expensive option is to use graphics processing units, but statistics show GPU’s are idling 98% of the time when processing the same 500GB of data (Trelles, 2011).

So we need to consider both cost and efficiency when choosing hardware for big data analytics.

When it comes to maintaining and improving system hardware designed for big data analytics there are generally two main approaches: vertical scaling and horizontal scaling.

Horizontal scaling means you add more machines onto the system; rather than upgrading the current ones. This means a company can just add more machines when they wish to process larger quantities of data and the system can be increased in small steps (Saeccker and Markl, 2013). The limitation of this approach is that the analysis software must handle the distribution.

Vertical scaling on the other hand means hardware is upgraded with faster components, such as the CPU and memory. This had the advantage of being to handle upgrades transparently and the analysis software will be able to adapt to the change (Saeccker and Markl, 2013). The disadvantage of vertical scaling is that the initial cost of upgrading the system can be expensive and the additional performance gains often go to waste until more strain is put on the system (Saeccker and Markl, 2013).

V. CONCLUSION

It wasn’t too long ago that a single gigabyte of data was considered to contain a vast amount of information (Bollier, 2010). Now we often talk about data in terms of terabytes.

In the future we expect petabytes and even exabytes of data to be processed (Trelles, 2011). This forces companies to make numerous considerations for how they are going to handle such large amounts of data.
Tools like Map Reduce are becoming more prominent in the industry due to their scalability, but we must also consider other technologies like parallel databases and programming languages design specifically for big data processing; such as R. Furthermore, we have to consider the potential security risks proposed by tools like map reduce.

In terms of hardware, we are expecting data-sets to become more high-dimensional in the future (Trelles, 2011), meaning a vertity of access nodes to the data, therefore system hardware that is both scalable and efficient will need consideration. We also must consider how we maintain the hardware and whether or not upgrading/buying new systems is worth it.

VI. REFERENCES


Big Data On A Small Scale

Using Big Data To Fight Showrooming

Adam charlton
Computer Science Student
Derby University
A.Charlton1@unimail.derby.ac.uk

Abstract—This article explores the methods that can be used in the retail world to combat the effect and use, of customers showrooming in which they use their smartphones to find a product online at a cheaper price after trying it out in a physical brick and mortar store. By customizing the experience for each individual customer, predicting trends, giving the ability to allow customers to find reviews or specifications on their products can remove the need to look elsewhere and secure the sale.

Index Terms—big data, retail, showrooming, personalized marketing, brick-and-mortar, e-commerce, facial recognition, NFC, QR, analytics, trends, markets.

I. THE FUTURE OF RETAIL IS ALL IN BIG DATA

Even before the term Big Data was popularized Walmart understood by harnessing the power of the data they collected through sales, spending habits and other factors, they could streamline its complex supply chain and thus limiting excess inventory and associated costs [Kelly, 2013]. These saving made were passed onto the customers in the form of lower prices, sometimes at considerably low discounted prices compared to its competition.

However since the early 2000s, many other retailers have pushed Big Data into their action plan, most noticeably Amazon and their innovative used of Big Data in which they looked at customers buying patterns and behavior, to give recommendations at the checkout based on any similar or related items that could be of use towards this sale.

The ever increasing amount of retail channels and social media have empowered customers, giving them access to a vast amount of information ready at their fingertips, being that one in every five people in the world now own a smartphone new techniques have arose for finding the best price around [Heggestuen, 2013]. Showrooming is the term is giving to the practice of a customer going into a typical brick and mortar high street store, testing or examining a product before looking online for a cheaper price elsewhere. Due to the overhead costs of running a high street store typically prices are far dearer than its online counterparts. Other than just losing the sale, these stores can having problems resulting in damages to the product floor samples available for browsing.

A survey conducted in 2013 showed that 40% of United Kingdom customers have used their mobile phones in the last year to find an item online for a better price [Moth, 2013]. It isn’t just the price point they are concerned with, they are also looking for customer reviews taking up between 15-20% of showrooming uses [Moth, 2013]. As well as 59% of shoppers looking up product information rather than talking to the in store advisors [Morran, 2013].

Retailers need to look at this trend and not ignore it, instead they should be looking at techniques that can be used to counteract the effects of showrooming, especially independent stores than cannot afford to lower their prices anymore.

By collecting, managing and analyzing enormous amounts of volume, variety, velocity and veracity of data, retailers can succeed in combating the challenges faced with customer’s showrooming. Using this data, retailers can generate valuable insights into personal marketing, optimizing their effectiveness of marketing strategies, decisions involving assortment and merchandising, and removing any inefficiencies in distribution and operations.

The information that is available continues to grow exponentially almost following the pattern of Moore’s law, in which the aggregate amount of data available to retailers doubles every two years [Woods, 2013]. Along with estimates of 90% of the world’s current data has been created in the last two years alone [SINTEF, 2013], however this data is not all associated with retail it is still a surprising amount. Data from social networking sites can also give an external insight into important trends that are happening right now too, even if they are not related to your paying customers.

Todd Hale, a CIO at Office Depot, says that with today’s analytics, “You are really shifting away from traditional report development into more of a data scientist role, where you understand the types of data available not just internally, but externally, and how to combine different data sets to answer questions that you couldn't before”.[Economist, 2014].

The retail industry is one that produces a massive amount of information from a range of sources, and one that understands the challenges and possibilities that arise from this massive collection of data. Though the single biggest barrier to effectively using this data, is trying to figure out what is useful among the excess of data. UK’s furniture retailer, DFS and its chairman, Richard Barker, a retail industry veteran, runs a loyalty card program for it business and explains the challenges they face with the vast amount of data they handle, “Data collection is now relatively easy in all different forms, but data is arriving in such a tsunami it’s hard to really make sense of it other than at a very common-sense level”.

According to research in 2013 a typical billion dollar retailer will spend on average $75,000 alone just on big data,
and only allocate less than half of its IT budget on new investments aimed at big data. However it is estimated that the investment cost will increase 1.6% by 2016, signifying a move towards big data applications and its value directed to their business operations [Lopez, 2013].

Though since 2012 the biggest challenge that faced retailers was managing the volume of data they received, in which it was 46% in 2012 has dropped to only 24% in 2013. Now previously where the volume was the problem, has now turned into the variety of data they are receiving, predominantly with unstructured data. Stating in the same report that now in 2013 68% of retailers rate the handling of unstructured data as their biggest challenge.

This article will explore some techniques that are being used in retail around the world, which take large quantities of data and turn it into useful informative findings that benefit, not only sales but the companies reputation on how it handles its customers.

Then onto how small and medium sized business (SMEs) can still use the same techniques regardless of its size, and the amount of data they can obtain through certain channels that have been made available to the public. As well solutions to setting up a Big Data rig and the options available.

II. THE UNIQUE SHOPPING EXPERIENCE

One advantage that traditional brick and mortar stores had over its online competitors, was the ability to purchase and obtain the product in the same day. Now we see the likes of Amazon, eBay and a range of other company’s providing similar services that allow same day delivery from its local stores [Parasuraman, 2013, Tsotsis, 2013], giving instant gratification to their customers and eliminating the last compelling differentiator that those high street stores had.

Does the need to visit these stores become redundant when giving the option to browse, order and receive the product in under 24 hours without having to leave your home. In fact no, a study by PwC unravels a selection of myths associated with multichannel retailing and online consumer behavior. The results show that for the majority of the categories found in retail, customers still favor going into a physical store, giving the option to browse, order and receive the product in under 24 hours without having to leave your home. In fact no, a study by PwC unravels a selection of myths associated with multichannel retailing and online consumer behavior. The results show that for the majority of the categories found in retail, customers still favor going into a physical stores [Montgomery, 2013] (see figure below). While many retailers employ a multi-channel strategy there is still an equal number of small and medium sized companies that are still working out from their physical storefronts only. Though these physical stores are important to the overall shopping experience, the key is now to personalize each unique encounter with the customer, in which they receive value and pleasure from visiting the store.

Big Data tools and techniques that have been successful in well-known e-commerce stores, have started to appear with similar capabilities in physical stores. These are:

A. Applying custom tailored experiences to individual shoppers.

Online stores have been successful in their delivery of content by personalizing it to their individual customers, they seem to have a more personal connection to them, by knowing their name, recent purchases and the ability to showcase relevant products all customized from their viewing habits and purchases. These forms of tailored made experiences have become to seep into physical store fronts.

For many years numerous retail stores have offered loyalty card programs and have been successful in marketing goods that customer might find interesting based on previous sales, brand loyalty and cross channel preferences.

Research conducted by the Harvard business blog showed that personalization can deliver between five and eight time on the return on investment used on marketing and increase overall sales by ten percent. Shoppers don’t mind sharing personal details as long as they see the rewards and benefits from it.

Though companies like Almax and NEC have envisioned an even more personal experience for retail stores, they have created systems that use facial recognition and cameras to analyze a shopper’s face and eyes, to detect their age, gender, ethnicity and further characteristics.

The face data that is collected is encrypted in real time and converted into characteristic data ready for analysis, there is no risk of accidentally disclosing images of the consumers faces as the characteristic data cannot be restored into its original form [Diginfo, 2012].

Along with the data accumulated from the camera, and the data and time in which that customer visited the store, these factors can be used to analyze trends in customer behavior and the frequency they visit the store. Also scenarios such as this could be a reality, in where after walking into your favorite retail store the cameras at the front door detect immediately who you are. The data that has already been collected on yourself will be sent to an in store advisor, through the retailers associate sale app. Displaying not only past purchases, and loyalty but providing a consolidated view of all your shopping characteristics. Using this information sales assistants could greet customers by name, ask about their previous purchase and enquire to see if there is any problems. Offering such a personalized experience can dramatically alter a customer’s view on the company, and with these latest advances in facial recognition and analytics that happen in real time, this reality could become the norm very soon.
B. Predicting Trends For The Latest Releases

Social media already plays an important role in retail, with many retailers using social networking sites as an easy connection point to its customers and to monitor reviews and feedback. Though by taking a step further and analysing people’s likes, posts, interests, and events they attend. They can create profile of popular items, even items such as games and movies that haven’t been released yet. The advantage of knowing what items are going to be popular can give the company the benefit of ordering enough stock to meet the demand and not miss out on any sales due to having no stock.

C. More Than Just The Price Tag

Many of the users that showroom only do so to gain more information about the product such as specifications, size, warranty and even if it comes in other colours. By having a means of displaying this information could be the once factor stopping a customer, looking up for the product by themselves and finding it cheaper at the same time too.

However displaying this type of information is not ideal in many situations due to the environments given and the space needed to be able to clearly present it. Although by using technology such as Near Field Communications (NFC) or QR codes, retailers can place these tap points next to the price tag allowing access to the information the customers need and want.

Purchasing a bike is considered quite a large purchase and many customers want to be certain they are making the right choice before spending a lot of cash. Evans Cycles, a leading bike showroom across the UK, has begun to improve it’s in store experience by providing all the relevant information through the use of QR codes. The QR codes link up to the appropriate product page featured online, were it displays full specifications, stock for both in store and online delivery, and reviews left by other customers who have already purchased that product [Qrcodepress, 2013].

Another UK leading online and high street store, Argos, has just implemented NFC capabilities in 40 of its stores allowing shoppers to tap their mobile phones, being that they meet the requirements, to engage with store staff who are wearing NFC enabled lanyards. Customers can find out the latest offers and recommendations, as well as being encouraged to install the Argos mobile app, increasing mobile sales and improving the multi-channel experience [Proxama, 2014].

These NFC and QR codes can used also used to profile customers into their handsets by recording what device was used, after that if they visit the website again at a later date using the same phone the system will recognise who it is and understand where their local store is. Giving the ability to push sales, events or any other useful information that could be relevant to their local store straight to their phone without the need for a consumer account.

This is only a selection of the things that are made possible from collecting and analysing patterns inside of Big Data, the volume of data will always be growing exponentially, it’s about how you handle it to produce new methods and techniques to take advantage over your competitors.

III. BIG DATA IS STILL AVAILABLE TO SMEs

The retail industry is becoming more aware of the potential that Big Data could revitalize the industry, which has been challenged by an increasing number of channels, a slow economy, and changes in consumer behavior online.

Being able to maximize the potential benefits of Big Data is difficult and requires the correct hardware, in-house knowledge and knowing which practices are the best.

On many occasions people have said SMEs cannot use Big Data as they don’t have enough of it. The EVP at SAP, Steven Lucas thinks differently, saying “Every company should be thinking about their big data strategy whether they are big or small”.

Although while it is true that SMEs don’t have access to the same volume of data found at the larger companies, even with smaller amounts of data, companies can still develop a Big Data strategy. There are plenty ways of increasing the amount of data you have for analysis though, public data sets are becoming more available for free, or for purchase and they can be combined with the existing data they already own to create complete new insights such as target groups or new markets.

Nike for example shares all of its data with the rest of the industry, containing data from all of their suppliers. In doing so allows for other organizations in the supply chain to use the data set to make better business decisions [BigData-Startups, 2013]. Finally SMEs should be looking for new ways to collect data themselves by sensors, loyalty systems and online accounts. However Big Data is a two part process, the collection of data and the analysis of data, even with the little data they might possess many any analytical techniques can be still used.

Indeed SMEs have to be more flexible as they have a limited amount of resources, nonetheless if you have the necessary skills in house there are open source tools available, in nearly field of data analysis including; date mining, business intelligence, data analysis platforms and the databases that hold everything together, just at the cost of hardware.

On the other hand cloud based solutions such as Causata or Ayasdi remove the need to build your own data rig, and can be scaled down to the amount you need reducing any excess cost. As well compared to owning a private platform, it can be very cost effective when the volume of data starts to increase.

After obtaining a data analysis platform and collecting the data, it’s all about how you can use it. The same techniques that the larger companies have been implementing using Big Data need to be applied to SMEs too.

IV. CONCLUSION

Data is becoming easier to obtain and through a complete new range of sources. Mobile phones create breadcrumbs trails of data containing the whereabouts of its owner using GPS positioning. The same happens when we browse the web and use a credit card, it all gets recorded. With the increase of data, the tools used to capture, collate, store and analyse it have also developed becoming better and faster.

With this new stream of technologies the challenges that SMEs face with Big Data is how they can use their data to
create new innovate and enhanced services that benefit the customers in not only in price, but with loyalty so they keep coming back.

Though predictions state that in 2017 there will be an estimated increase of 243% for the demand of Big Data specialists [is4profit, 2013]. We could see start-up companies specialising in the field just to be commissioned by the smaller sized businesses, to sort and analyse pre-existing data.

Big Data can be overwhelming especially to SMEs, they must understand the limits of what their current systems can handle, and for them to produce results retailers they must act quickly on any insights gained. As well staying up to date in both human knowledge and technology power. However a lot can be gained from the information you gather and analyze, possibilities are endless.

V. REFERENCES


Big Data In The Music Industry
Is The Risk Worth The Reward?

Alex Clark
Faculty of Business, Computing and Law
University of Derby
Derby
100175836
A.Clark11@unimail.derby.ac.uk

Abstract—Big data has bred a new way of analyzing data faster and bred a whole new industry in IT with companies clambering to take advantage of it. The music industry is no exception, small and medium size enterprises in the music industry such as record labels and artists want to know everything they can. However what they don’t know is that big data can be extremely unreliable. If they are aware of how unreliable big data is and the potential risks that come with using it, they can use it more responsibly and maximize its benefits.

Index Terms—Music industry, big data, reliability, risks.

I. INTRODUCTION
With big data set to be one of the next big things in technology, there is a mad rush in business to understand and extract value from it. The music industry has scrambled to take advantage of this with a recent deal between twitter and a company called 300 allowing them access to tweets dating back years (Oswinski, 2014). As well as the UK government announcing that part of its seventy three million pound investment in big data will be a project which will benefit musicians (Gov, 2014). This will give the company a new insight into the music industry which they can attempt to exploit for business purposes.

Potentially big data and the music industry could work very well together. It could give all sections of the music industry, artists, record labels, promoters and streaming services a whole new insight which they never had before. It could give the whole industry a chance to revolutionize itself. It could change the artists that fellow artists tour with, give potential to new music styles and target bands to people who are more likely to like them and generally make the music industry a better and more profitable industry. However big data has the potential to be very risky.

II. THE RISKS
A. Reliability Risk
The first issue surrounding big data is the issue of the age of any data collected. Data about bands and music goes back hundreds of years. However, anything older than two years is seen as irrelevant because it is outdated. That makes anything beyond roughly two albums old for an active band useless. This will heavily limit the impact that big data can have for the music industry. There is also the risk that on top of that anywhere from 0.1% to 5% of data will actually be useable (Wu, 2012) and if this is analyzed incorrectly you could end up with barely any relevant data.

The next issue is highly important in all sectors for big data and in particular musicians which is the reliability of social media. All bands have social media pages. They use them to advertise everything they are doing, sell merchandise, promote tours and generally keep in contact with all their fans. Their comments and opinions can change everything that the artist does. Such as where they go on tour or the style of music they create. Big data could analyse everything on social media and show the best places to tour, as well as potentially who to tour with in order to maximize ticket sales.

However social media data is really unreliable. It is estimated that there are around eighty three million fake facebook profiles (Kelly, 2012) as well as it being estimated that five percent of twitter accounts are fake. It has also been estimated that one third of people lie on social media (Warman, 2013). There are many reasons for lying and although it is highly unlikely to lie about artists you like it is always possible. The question about social media remains though, how reliable is any data that you receive? One recent estimate stated that around 80% of social media data could be unreliable by 2015. Currently 12 terabytes of tweet data and 25 terabytes of facebook logs are created every day (Easton, 2014), which is a massive amount of potentially unreliable and risky data.

Alongside social media data there are massive sets of potential data from streaming services such as Spotify and YouTube. However this also comes with risks, the main risk is the misinterpretation of the numbers. YouTube has a comments section, likes, shares, dislikes and views data available on each page all of which are valuable pieces of information to the music industry. What big data can’t show is how many times a person has viewed the same video. Also thumbs up/down doesn’t show the true popularity of any video many people watch videos and simply stop if they don’t like it. It is believed that YouTube views don’t count unless the entire video is watched. Also, because with information in the music industry such as social media comments and YouTube views you become more likely to misinterpret data when there are more variables (Taleb, 2013). The data that is so valuable then becomes very unreliable.
The final issue that shows the unreliability of big data is that to a certain extent in the music industry it isn’t really needed. Record sales, merchandise sales and the ability to sell out shows are a more reliable source of information to show you how popular a band/artist is, not lots of data which may be years old.

B. Legal Risk

Reliability is not the only type of risk associated to big data, with big data there are many potential legal issues. A lot of big data that will be relevant will come from music streaming sites such as YouTube, Spotify and social media sites such as Facebook and Twitter. This could easily fall foul of data protection laws and privacy laws. There has been a long running debate about whether anything said on social media is private (Syme, 2013) and advertisers have used that to their advantage. Social media can be seen as a place of free speech but the reality is that once people know that it might be used to target them privacy laws will protect them.

The biggest legal risk is the Data Protection Act (1998). With lots of data that the music industry will be interested in being on social media the biggest risk is not sticking to the law regarding anonymity. This is a massive risk and even the police have been found to break this law, with around 1100 offices to have been found breaching it from 2008 to 2011 (Condon, 2011). This can be managed with a solid data protection policy but even they don’t guarantee that companies won’t breach this law. Recent high profile cases such as the NSA leaked documents show how easily data that is supposed to be protected can easily be leaked (Whittaker, 2014).

It is true that data that is anonymised is protected under the Data Protection Act (1998). But this has two real risks, the first risk is that even if you think the data is anonymised there is a risk that a mistake could have been made and you are liable to be prosecuted (Walker, 2013) which in one case cost a six million dollars. The second risk is that when data is anonymized there is very little you could do with the data which would be beneficial (Walton, 2014).

C. Reputation Risk

There is not only a legal risk in this, there is also a massive reputational risk as if it’s known that the data is being exploited and it goes wrong, for example when Target were predicting pregnancies, the fall out can be catastrophic as users will be much more careful with their data round the company and could affect sales (Bhasin, 2013).

A lot can be found out through big data and with social media apps such as BandsInTown you could easily start to predict who is more likely to go to shows and attempt to directly to market to them more. You could then also potentially see who is more likely to buy merchandise through photos on social media pages and again try to directly market to them. These are great features of big data but the Target pregnancy scandal showed that there is the potential to get carried away (Hill, 2012).

The reputational risk to a certain extent is more important than and legal or reliability risk. Two fifths of people affected by data failures say they suffered reputational damage as an additional result of any failure (Global Legal Post, 2013). If a record company is found to be exploiting customer’s data without their knowledge this not only violates the Data Protection Act (1998). Customers could easily stop using their services, in the case of the music industry this could include not downloading any records by that music label.

In order to counter the risk most organizations have some sort of risk management team to try and minimize the risk which comes at a heavy cost (Davenport, 2014), so there are financial implications in the use of big data which SME’s in the music industry may not be able to afford.

D. Misuse Risk

Big data can show you a lot about your customers and potential customers and it can be very hard to draw the line between what you need to know, what you want to know and what you don’t need to know. The distinction between those can be blurred and it can be easy to forget that you’re dealing with people (McDonald, 2014) and those people are the heart of the music industry. The fans and the artists are all people and they can easily walk away.

There is also the risk that big data could show connections that are there which appear to be linked but in reality aren’t. Big data isn’t an exact science and is in its experimental stage as a form of IT and there isn’t an exact way to see what you want to see which can unfortunately mean that it is misinterpreted (Osak, 2014).

There is also the risk with the misuse of big data that claim that big data can solve all your business problems and give you new insight into your customers, the reality is that it can’t. The problem is that people easily believe the hype of big data without understanding the real limitations of it. One survey showed that 53% of people believe big data will have a positive impact (Anderson & Raine, 2012). But this is due to lots of positive stories on big data and not actual evidence that big data will have a positive impact. Data can also easily be skewed or falsified and big data can’t check that, it just shows you what is available (Clark, 2013).

There is also a belief that big data can help predict future trends (Chemi, 2014). Shazam have recently used big data to predict next year’s big artists (Datoo, 2013). The reality is that to do this you would need real time analytics which can cost $100,000 per 1 trillion GB’s of data (Cohan, 2014) and all predictive analytics is, is educated guessing so the reality is that it could fail horribly. Also predictive analytics is based on using data in the past and present and assuming that it will to a certain extent repeat itself. With the music industry constantly changing it is highly unlikely that predictive analytics will provide any real benefit.

E. Financial Risk

With all the other types of risk there is also a huge financial risk. In 2011 1.8 trillion gigabytes of data was created (Tobin, 2013) and it has expanded ever since. However in order for big data to be used there needs to be a place to store the data acquired. Which means additional money has to be spent on servers or data centers. Data is cheap but the reality is that setting up the operation behind data centers or servers isn’t.
Along with that you need a team of data analysts and some sort of analytical software all of which costs lots of money. So in order for the financial risk to be justified there needs to be clear aims for the data use and some evidence that the data can deliver results (Weathington, 2014). One recent estimate is that 57% of people involved with big data are worried about how to extract value from their data (Jordan, 2013). At the moment there is a range of different objectives with data in music and many other industries but no real proof that it will be successful as its still in an experimental stage (Robinson, 2013). The irony of big data with financial risks is that 37% of people who are currently using big data are using it in the hope of reducing costs (Jordan, 2013). The large music organizations can afford to take this financial risk but for SME’s in music such as artists and small record labels the effect could be catastrophic.

F. Security Risk

Along with the other types of risk big data has a huge security risk. Big data although bought cheaply is valuable to hackers. Recently one of the biggest entertainment based companies Sony was hacked. The target for the hackers was the data about its gaming customers, but the cost of the damage was both financial and reputational (BBC, 2013).

The problem with big data is it requires extra computing hardware to facilitate naturally making it a more valuable target to hackers, because not only would they be interested in the data, but any data associated with it such as current customer’s details would make it a more valuable target than other organisations. It’s also more likely that SME’s such as small record labels security won’t be as good as large organisations such as Sony (Wee, 2014). In order to accommodate for the increased potential threat physical and cyber security and security protocols would have to be updated so that they don’t become so vulnerable to threats.

With all computers there is the legal threat of the computer misuse act. The main problem with this law is that it ‘prohibits unauthorised access to, or modification of, computer material. For the purposes of the CMA, access is unauthorised where it is not consented to by the person entitled to control access.’ (Rudgard & Bentham, 2013) This is important for big data because in the debate about whether social media data belongs to the user or the company it opens the possibility that anyone who uses data could be prosecuted under the computer misuse act even if they think everything is legal.

III. POTENTIAL ALTERNATIVE SOLUTIONS

A. Use data available

The biggest problem seems to be that we as organizations don’t properly use the data that is already available (Ross et al, 2013). Big data may show you where exactly people listen to a band’s music worldwide but that can easily be done by other methods. For example checking ticket sales for the last time the band toured there. Or checking social media and simply looking at where the comments are coming from. This is a free way of using data available and wouldn’t require a lot of new software, staff and other pieces of IT. All this would require is a bit of time to understand the customers better.

This could also give you an insight without using big data. For example if you wanted to know how many fans you had got after a tour you could simply check the amount of Facebook likes on your page before and after a tour.

B. Ask The Customer

Another simple free alternative is one which seems to be commonly ignored which is simply asking the customer (Aaronson, 2013). Bands and record labels have easy access to their customers, they should utilise this by simply asking their customers of their opinions and views. This is a simple free form of data collection and that way they are more likely to maximise customer satisfaction. In one recent case it has worked for the band Mallory Knox who asked their fans where they wanted to play and who support them and the tour sold out all but two of its dates.

There seems to be a common business practice to not ask the customers because only people who are angry respond (Mates, 2013). However music is different to other industries and generally people are more likely to be honest and help you improve.

C. Spending Alternatives

With all the risks that come with big data the reality is that big data isn’t reliable or worthwhile. Big data comes with some many risks that in reality it isn’t beneficial to the music industry. In total around three trillion dollars has been wasted on IT failures annually (Krigsman, 2012). The music industry doesn’t have lots of money unless they are the big record labels such as Sony or big well known artists such as Arctic Monkeys or One Direction and they wouldn’t have any use for big data. The reality is that the music industry can be successful without big data. Lots of bands struggle financially and fail to pick up fans and sell out tours, but that’s part of the process. Very few bands are successful and with the short span of a lot of bands big data won’t give the bands the short term success they need and the record companies the additional money from the bands increased popularity.

In a world of illegal downloads financially the music has taken a massive hit. Around 95% of music downloaded online is illegal and the estimated loss each year is around 12.5 billion dollars (Go Gulf, 2011). But spending money on big data isn’t the solution for increasing revenue. The reality is that legal and reliability issues make big data a project which in all likelihood would do more damage than good. An alternative solution to this would be to spend any money trying to shut down illegal sites and generally trying to discourage illegal downloads. Streaming sites are hugely beneficial for the music industry as it allows potential customers to essentially sample music and gather fans (Sydell, 2012). But trying to help shut down the sites which allow music to be illegally downloaded would be more beneficial.

D. Improved Advertising

Small club show attendances are down (Greenburg, 2012) but festival attendances are up (Laing, 2012). That will help to reverse the financial impact and may help boost legitimate music downloads and with attendances falling it’s important to
reverse that trend (Owenskin, 2014). A stronger presence on sites which advertise music such as Facebook and YouTube would be more beneficial than big data. It is impossible to target a particular area even if big data showed that there was either a strong fan presence or a weak fan presence in an area.

However if YouTube could be convinced to started linking a band’s video from another the fan base may develop faster and would be a better use of money. Or as recent stories have shown there is the potential to pay for tweets and Facebook posts which could increase a band’s popularity through sites although that comes with about the same level of reputational risk as the use of big data.

IV. CONCLUSION

Big data is seen as the next big thing and many different industries are jumping onto it without any real understanding of the risks involved. It is seen as an effective way to make more money and understand customer’s better but very little is said about the risks. There is an argument that there is risks in everything in life and we won’t know until we try, but much like the dot com bubble it has the potential to burst with a huge amount of financial loss. The music industry should stay away from this as the risks far outweigh any potential benefits and if big data is a solution to a lot of IT problems then it might be worth looking at in five to ten years’ time.

V. REFERENCES


The Big Data As An Opportunity For Tourism Industry

Marcin Tomasz Drozdz
School of Computing and Mathematics
University of Derby
Derby, United Kingdom
M.Drozdz1@unimail.derby.ac.uk

Abstract—The purpose of this paper to identify and describe the main opportunities and benefits of Big Data implementation in tourism industry SMEs. It will also indicate how small and medium enterprises can gain information about their customers activity and preferences, process and analyze information. This paper contains information about challenges that SMEs can face during implementation process and ways of solving these issues.

Keywords - Big Data; SMEs; Travel; Tourism Industry; Data Analytics.

I. INTRODUCTION

Big Data is now one of the main fields of interest for many businesses, regardless of the sector of activity. “Big data is a collection of data from traditional and digital sources inside and outside your company that represents a source for ongoing discovery and analysis” (Arthur, 2013).

The use of a huge amount of unstructured data facilitates a process of decision-making, facilitates the creation of more innovative products and services and allows close relationships with customers.

“Big Data is about the growing challenge that organizations face as they deal with large and fast-growing sources of data or information that also present a complex range of analysis and use problems” (Villars, 2011).

In the UK, research on Big Data has been recognized as one of the eight strategic directions of development of the British sector of new technologies, next to the: space industry, autonomous systems, synthetic biology, regenerative medicine, agro-technology, new materials and alternative energy sources. (DBIS, 2013)

II. BIG DATA IN TOURISM INDUSTRY

A. Growing volume of data

Travel and Tourism Industry generates nine per cent of global Gross Domestic Product (TravelMail, 2013).

“In 2012 British tourists spent over $52.3 billion on holidays and made over 55 million trips abroad” (UNWTO, 2013).

“Tourism is one of the sectors with a very close relationship with the new information and communication technologies” (Miguens, 2008).

Over 50% of tourists use social media for travel inspirations, 76% of them post vacation photos to a social network. Almost 50% of travelers revives hotels online, 40% reviews holidays attractions and hotels (Piombino, 2012)

Social media and internet activity of travelers creates huge amount of useful and significant data. Those information can be easily used by travel agencies and travel industry through the use of Big Data.

B. Big Data In Tourism Industry

IT technologies undoubtedly affected the operation of the tourism industry. The most popular trends in computerization of tourism industry are: Big Data, cloud computing, mobility and social networking. The growth of interest in these technologies occurred due to: need of instant access to information, regardless of time and place, a huge number of data placed on the Internet and need of constant access to information and social communication.

Big Data works well in the tourist industry, affecting the future of travel and bringing potential benefits to both travel companies and travelers. Travel companies gain a large amount of data that should be properly processed in order to provide information about consumer preferences and for marketing purposes.

C. Data sources

Data sources usually come from inside of the company. For banks, for example, such sources are all our activities related to your account, or how, when, and for what we pay, or make some purchases on a regular basis, how much money is on the account, etc.

These data come from various sources, such as: the purchase/sale transaction, consumer activity on social networks, meteorological sensors, digital photos and videos, GPS signals from cell phones, public databases (McAfee, 2012).

A lot of companies collect the data on the activity of customers (with their permission) by the mobile applications. Big Data gives great opportunities, such as: faster, more accurate analysis capabilities and ability of data usage from multiple sources.

In the old model, when the data were difficult to obtain and expensive to process, companies need to decide what data they need to determine the structure of the database system before they started to collect data. Now the collection, analysis and storage of data is much more affordable. The falling cost of
data collection and processing makes us able to solve computational tasks that yesterday were not economically justified or even seemed impossible (Davenport, 2012).

III. BIG DATA AS AN OPPORTUNITY FOR TOURISM INDUSTRY

Tourism market expects quick and easy access to services and information via mobile devices, such as smartphones, tablets and notebooks. Tourists want to have control over the planning and execution of a trip at any time, using a mobile device with Internet access. Mobile devices are also useful in business travel, used to change a reservation, check the weather or find the nearest ATM. Mobile application market will grow by offering travelers more options of trip planning and booking. That’s a great opportunity and field to develop, especially for SMEs. In the coming years, tourism companies will be looking for the Big Data solutions to resources planning, tariffs and prices management, support of additional services, customer service and business trips.

A. Opportunities for SMEs

Small and medium enterprises are still looking for solutions, modern technology that enable the management of operations and sales through multiple channels, while offering easy access to the required information.

Big Data gives new opportunities for the tourism industry, such as: (Davenport, 2013).

- Better support of decision making
- New products and services
- Improved customer relations
- Cheaper and faster data processing

B. Characteristics of data in the travel industry

Volume, Variety and Velocity are the main properties defining the Big Data. For the travel agencies internal sources such as costumers databases are primary source of information. Costumers activity on the internet, in social media creates variety but important information which should be structured and processed for marketing and planning purposes.

<table>
<thead>
<tr>
<th>Big Data feature</th>
<th>Characteristics of the tourism industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>Internal sources: Customers databases, transaction history, online ticket booking history etc. External sources: Large amount of data posted on social media, users reviews posted on recommending websites etc.</td>
</tr>
<tr>
<td>Variety</td>
<td>Large variety of external sources: data posted on social media, on the internet</td>
</tr>
<tr>
<td>Velocity</td>
<td>Variables preferences of travelers, price volatility, market volatility, activity of the competition, market seasonality</td>
</tr>
</tbody>
</table>

Fig. 1. Characteristics of Big Data (According to the Tourism Industry)

IV. CHALLENGES

Internet and social media activity of tourists creates a large amount of data that can be used by Tourism Industry to increase the flexibility of traveling, concentrate travel agencies efforts on needs and preferences of passengers.

Tourism industry does not use the full potential of the available data. “Most companies are far from accessing all the available data. Some haven’t even mastered the technologies needed to capture and analyze the valuable information they can access” (Bughin, 2012).

For the tourism industry Big Data is not only a chance for development, but also huge challenge. Tourism companies face technical and operational challenges associated with the use of Big Data in the industry, such as:

- The creation of an integrated data sources
- Working in the technological environment
- Lack of staff specialized in the field of data processing
- Keeping ahead of the competition

“Data analysis is considerably more challenging than simply locating, identifying, understanding, and citing data. For effective large-scale analysis all of this has to happen in a completely automated manner. This requires differences in data structure and semantics to be expressed in forms that are computer understandable, and then robotically resolvable” (Agrawal, 2011)

One of the major problems is the fragmentation of key data and their distribution in the different units and departments of travel agencies. In the typical medium size enterprise, information are divided between the customer database, information related to marketing, such as customer preferences, loyalty programs and external sources, such as social media. Implementation of an integrated database is expensive, time consuming and it requires extensive experience (Rijmenman, 2013).

If the tourism agency wants to use Big Data, it must face the unstructured and various forms of data, that continuously flows. It also has to transform unstructured information into a form which can be subjected to the tests, and then to analyze them continuously.

“All this data must be molded into an information foundation that is integrated, consistent and trustworthy, which were the leading data priorities cited by our respondents. Therefore, even though smart organizations will start down the analytics path by selectively attacking the biggest problems (and selectively building out the parts of the data foundation most relevant to gaining insight about them), they’ll need to understand how each piece of this data foundation aligns to an overall information agenda” (LaValle, 2011).

Big Data systems should interact with the existing hardware, software and databases, that’s another problem for developing companies. Previous versions of IT software, devices and databases will be still necessary and useful to analyze and optimize business activity. After the implication of Big Data solutions, some of SMEs will have to work in hybrid environment - using old and new version of software, at the same time, during the implementation presses.
The final and most important challenge for companies in the tourism industry, is to maintain a sustainable competitive advantage through the use Big Data.

Data resources are still not properly used by the tourism businesses. SMEs should gather the necessary skills and develop an overall strategy for the use of databases. The next step is to imply the Big Data solution as soon as it is ready.

V. SOLUTIONS

A. Big Data Processing In Cloud Computing Environments

“Cloud computing is an extremely successful paradigm of service oriented computing, and has revolutionized the way computing infrastructure is abstracted and used” (Agrawal, 2011).

During last few years cloud computing has become one of the main solutions for data processing. More and more of specialized companies provides big data cloud computing solutions. (McKendrick, 2014).

Cloud computing enables SMEs to implement big data solutions without investment in hardware. “The rise of cloud computing and real-time data processing is enabling retailers to target offers at their customers far more accurately” (Wall, 2014).

VI. BENEFITS

A. Better Support In The Decision Making Process

A lot of travel companies use Big Data solutions to make better operational decisions. Access to large databases gives the possibility of improving the possess of customer focused decisions making (better meeting customers’ expectations, holidays offers, destinations) (Regalado, 2014).

B. Development Of New Products And Services

Possibility of new products and services developing is one of most important features of Big Data solutions. Big Data provides an information about the personal preferences of customers, inconvenienced they face in the process of travel booking or while traveling, and many other valuable information that may contribute to the development of new tourism products (Vivion, 2013)

“Personalization is a key tenet of Big Data. With so much available information about a particular consumer, transaction or destination, the reality is one of detailed, accurate personalization. In order to most effectively win at true personalization, travel companies work across silos to gather the myriad data points created by a consumer at different points (Vivion, 2013).

Big Data can have a significant impact on the quality of travel, it will provide benefits for tourism businesses and travelers.

C. Cheaper And Faster Data Processing

New IT products, services and solutions offers better price-performance ratio. The large amount of data that must processed in the tourism sector makes it necessary to implement cheaper and faster technology to support data analysis. Cloud computing and open source software creates a new possibilities of development especially for SMEs (Burns, 2013).

D. The Revenue Management

In the tourism industry, optimization of products and services, such as booking of an airplane or hotel rooms, is one of the most important fields of planning and analysis. “On the operations side, using big data can improve the cost-benefit ratio of travel an average of 10% to 15%” (Gebhart, 2013).

The use of Big Data contributes to a significant increase in revenues and profit margins through an integrated revenue management system, that can be able, for example: to estimate the likelihood of cancellation or failure to appear of passenger for check-in at airport (Davenport, 2013).

E. Other Benefits

Big data solutions used in the tourism industry has also a positive effect on: (Davenport, 2013)

- Optimization of costs and returns on investment
- Personalized service
- The transformation of business travel
• Optimization of internal operations
• Improving financial results

VII. CONCLUSION

The concept of Big Data is already having significant impact on key processes in the tourism industry. Some sectors of industry and large companies are already testing, implementing or using Big Data, while many others haven’t take the first step in this direction yet.

Big Data creates new fields for development of planning, data analyzing, financial management, creating new, better offers and services. For SMEs Big Data is not just an opportunity, it’s also a huge challenge. To face that challenge companies should concentrate their efforts in particular areas, such as: acquisition of knowledge and experience in the field of big data and establishment and implementation of strategy for Big Data, procedures for data analysis, determine the sources of information, determine the most important goals for company. Big Data implication in tourism industry can result in optimization of costs and internal operations, it can improve financial results by providing new planning and analytics capabilities, that are essential especially in the field of product and service personalization.

VIII. REFERENCES


Abstract—Big data is a rapidly growing area of ICT that is becoming ever more vital to competitive industries such as retail. Failing to adopt and benefit from big data analytics could force a business to be out-competed in a competitive environment. There is however an alarming shortage of data-scientists who have the knowledge and experience to unlock the benefits of big data. This shortage is also predicted to grow towards the end of the decade.

This paper explores some of the opportunities that big data has to offer, the challenges involved with recruiting data scientists, and what can be done to make the best out of a challenging situation.

Index Terms—Big Data, Retail SMEs, Data Science, Data-Driven Decision Making.

I. INTRODUCTION

In the second decade of the twenty first century, the competitive nature of business is putting pressure on small and medium sized enterprises (SMEs) to use everything in their power to maintain a competitive edge. Research has shown that over half of all new businesses fail within the first year (Ucbasaran et al. 2013), and one of the most important areas for gaining a competitive edge is the adoption of Big Data and data-driven decision making techniques (Manyika et al. 2011).

Big data, which is still very much in a developing stage analogous to the web 1.0 era (Provost & Fawckett 2013), is a rapidly growing area. Research conducted by The McKinsey Global Institute (Manyika et al. 2011) has shown that big data can increase an organizations operating margin by up to 60%. A study conducted by Tambe (2012) at the New York University Stern School also showed that even a slight increase in big data had a significant increase of productivity.

While this sounds too good to be true, adopting and making efficient use of big data comes with its challenges.

Firstly, as data becomes cheaper and easier to obtain (McAfee & Brynjolfsson 2012) there is a growing demand for data scientists that is not being met by supply (Davenport & Patil 2012). This can be problematic, not only because of the initial shortage of experts, but also because of the use of employees that do not have the correct skillset for data science can be harmful to a business (Shah et al. 2012).

Secondly, the amount of data generated globally is growing at an exponential rate, with sensors, computers, and mobile phones generating ever higher quantities of data (The Economist 2010). This is pushing big data into a new era which is still in its infancy (Provost & Fawckett 2013), meaning there is a shortage of knowledge and experience in many areas of big data.

If a retail business is to benefit from this growing area of ICT, it is important to consider the challenges involved with incorporating big data into complex decision making, and avoid pitfalls that can be potentially harmful to a business (Shah et al. 2012).

II. BIG DATA – OPPORTUNITIES AND BENEFITS FOR RETAIL SME’S

A. The Benefits of Using Big Data

Technologies such as smartphones, the internet, computers, sensors, smart energy meters, scientific research, and social networks, are generating ever increasing quantities of data that equates roughly to a ten-fold increase every 5 years (The Economist 2010). As of 2012, up to 2.5 exabytes of data is generated each day, and more data traverses the internet every second than made up the entire internet two decades ago (McAfee & Brynjolfsson 2012).

This drastic increase in data is driving the evolution of big data technology which has provided opportunities for large organizations such as Amazon and Wal-Mart to significantly increase their operating margins through big data analytics (Provost & Fawckett 2013).

A classic example of this is Wal-Mart, who used big data techniques to detect an increase in sales of specific items such as beer and Pop-Tarts in areas of the country that had a hurricane forecast. This information was used the next time round to ensure that these items were stocked up in the areas where hurricane Frances was forecast to hit, ensuring that demand was met (Provost & Fawckett 2013).

While this provides an example of big data put to use within retail, it is by no means the limit of what big data has to offer. Big data is still in a developing stage that is analogous to web 1.0. Once the infrastructure, software, and techniques have been developed further, big data will reach a 2.0 phase where businesses will be required to take full advantage of the potential benefits that can be gained from big data and data-driven decision making (Provost & Fawckett 2013).

B. Big Data Opportunities

While big data is being adopted in many different sectors, it can provide significant advantages to retail organisations that
exist in a highly competitive environment (Chen et al. 2012). Provost & Fawcett (2013) describe two major ways in which big data can provide opportunities for a business:

**Data-Driven Decision Making (DDDM)** can be used to support business decision making, instead of relying on intuition/experience alone. Using data to support decision making has been shown to provide significant advantages. A study conducted by Brynjolfsson et al. (2011) showed that firms which were more data-driven were also more productive, had a higher market value, and had a greater return on assets. While DDDM has been used for decades, big data provides an opportunity to push DDDM to a higher level than previously realized (Russom 2011).

Big Data analysis can help to identify new information including patterns and correlations within sales data, such as the previously shown Wal-Mart example (Provost & Fawcett 2013). This can also help businesses to acquire new insights into previously unknown business opportunities that would otherwise go unnoticed.

The vast quantities of data generated every year provide enormous opportunities for retail businesses to gain from big data analytics. The McKinsey Global Institute (Manyika et al. 2011) predicts a 40% annual rise in generated data.

As well as generating data through customer transactions, loyalty cards, and sales data, retail businesses can also purchase large quantities of data from external sources. Along with this, big data tools such as Hadoop, Hbase, and CouchDB are providing SMEs the opportunity to take advantage of the benefits from big data analytics (Provost & Fawcett 2013).

### III. CHALLENGES WITH INTEGRATING BIG DATA INTO A RETAIL BUSINESS

**A. Big Data is a Complex Area**

Utilising Big Data for tangible business benefits is not as simple as acquiring some data and looking for correlations within it. Firstly, the “Three V’s” that separate big data from other data analytics (McAfee & Brynjolfsson 2012) describe big data as having a high volume, variety, and velocity of data that provide many different challenges (Russom 2011). If a retail business is to maintain a competitive advantage through big data, it needs to manage a large, real-time, constantly updating, and varied data source (Buhl et al. 2013).

Drawing back on the previous example, Wal-Mart do not simply look for correlations within a database – in order to benefit from big data they have to manage an astonishing 2.5 petabytes of data generated from over one million customer transactions every single hour (The Economist 2010).

Similarly, if a business that handles transactions online, such as an online auction, there is often a need to process a very high throughput, and business decisions need to be made within milliseconds (Provost & Fawcett 2013).

There are also challenges with ethical factors such as intellectual property rights, confidentiality, and privacy when dealing with big data (Richards & King 2014). Ensuring the protection of privacy and data security is becoming more of a challenge as increasing amounts of data are generated and shared globally (The Economist 2010).

Because of these and many other challenges, it is essential to ensure that data scientists with the right experience and skillset are recruited.

**B. Recruiting the Right Data Scientists**

Unlocking the value held within big data is not a task that can be easily completed by anyone who learns to use big data analytics tools. Most organisations are run by people who do not have any idea about how to unlock the real value held within big data (Manyika et al. 2011). This is problematic not only because the full potential might not be unlocked, but also because in a highly competitive industry like retail, those who fall behind will be out-competed by those who manage to make more effective use of big data (Manyika et al. 2011). From this perspective, it is safe to say that recruiting big data employees who do not have the required skills, experience, and domain knowledge, could create a competitive disadvantage though bad decision making and inefficient use of data that may have been expensive to collect (Shah et al. 2012).

Moreover, in order to make effective use of big data, and add value to a business, data scientists are also required to have “domain knowledge”, with a deep understanding of business needs in order to ensure that the correct questions are asked, and useful answers are provided (Davenport & Patil 2012).

**C. A Shortage of Data-Scientists**

As big data is currently in a developing stage, and is a rapidly growing area if ICT, there is a very alarming shortage of available data scientists that possess the required skills and experience to unlock the full benefits of big data (Davenport et al. 2012). The gap between supply and demand for data scientists is also predicted to grow. Research conducted by The McKinsey Global Institute (Manyika et al. 2011) shows the US will need up to 190,000 more deep analytical talent positions and up to 1.5 million extra data-savvy managers in order to meet the demands of big data analytics by 2018 (see figure 1.)

![Fig. 1. The McKinsey Global Institute Data Scientist Shortage 2018 Prediction](image)

Adding to the issue, although data science is described by many as a popular growing area of ICT, there is a severe lack in university programs and educational training that specialize in data science and big data analytics (Dumbill et al. 2013).
This is only likely to postpone the gap between supply and demand into the next decade.

Another reason for the shortage of big data experts is the required attributes that are needed. Davenport & Patil (2012) describe a good data scientist as a “hacker, analyst, communicator, and trusted adviser”. The author describes this combination as being extremely powerful, and extremely rare.

Given the current high competition for data scientists, Davenport & Patil (2012) argue that it can be extremely difficult and expensive to find and retain data scientists that have a combination of big data analytical skills, experience, and the scientific background that is needed. Because of this, and the risks involved with using inexperienced employees, it is essential to carefully manage the process of recruiting and retaining the required skills and talent that a business needs for big data analytics.

IV. BIG DATA OPERATIONS – HOW TO DEAL WITH A SHORTAGE OF DATA SCIENTISTS

The predicted global shortage of data scientists is an issue that cannot be avoided. The best course of action that a retail SME can do is to plan and prepare for how best to deal with future shortages. The following sections give advice on what can be done to ensure that a business prepares for this event in the future.

A. Recruiting the Best Data Scientists

One of the more immediate challenges involved is ensuring that the correct data scientists are recruited. As previously mentioned, using employees who do not have the required skillset and experience can be damaging to a business. Having data scientists with the correct mindset is just as important as having access to big data analysis tools such as Hadoop, Hbase, CouchDB or others popular frameworks (Davenport & Patil 2012).

Previous experience and good skills with big data analytics is not the only requirement. Given that every business has different needs, it is important that a data scientist is flexible, and can easily adapt to the tools that are already integrated into a business, even if they have no previous experience with them. Not many big data analytics tools are tailored to the needs of any one business (Provost & Fawkett 2013), and it is important to be flexible.

Given the current shortage of data scientists, and the predicted growth of the shortage, one possible alternative is to look outside of the big data analytics field for the right talent. Davenport & Patil (2012) argue that many related fields, such as experimental physicists or the social sciences often deal with large quantities of data, and could provide alternative benefits from other angles. Looking for the right talent in related fields could prove to be fruitful for these reasons.

B. Integrating Data Scientists into the Business

There are limitations with regards to recruiting external experts, as well as training regular employees in data analytics. When a new big data analytics approach is adopted, it is typical for an organisation to recruit a team of experts in the hope that the required skills and expertise will trickle down the ranks and be passed on to other employees who will be able to fill data analytics roles in the future. In reality, many organisations remain stuck in the ‘expert’ phase, with the crucial analytical skills remaining concentrated within a small data-science team that are externally sourced (Shah et al. 2012). This separation of data-scientists from the rest of an organisation is damaging in the respect that it can create a lack of understanding within the data science team, as they are not fully aware of the business needs, and lack ‘domain knowledge’. This can be especially damaging when important business decisions are dependent upon big data (Provost & Fawkett 2013). Integrating the data science team into the rest of the business will also help to pass on skills and experience to other employees, who themselves could prove to be a valuable asset in the future (Provost & Fawkett 2013).

C. Bigger is Not Always Better

Another important consideration is the size of big data analytics that will meet business requirements. Big data analytics can be a costly and inefficient process, and bigger gains can sometimes come from analysing smaller amounts of data.

Useful sources of data can often be found in unexpected places. For this reason, intuitive and original methods for obtaining and analysing smaller quantities of data can be more effective and faster than using well established but slower methods of trawling though large quantities of data. McAfee & Brynjolfsson (2012) put this idea to the test by using publicly available web search data to predict changes in property values across the United States. This produced results with a greater accuracy than the official National Association of Realtors, proving that a small team of experts can product results that are more effective than larger data-analysis operations.

It is important to consider what business questions can be answered through big data analytics, and decide upon the most efficient and effective approach for unlocking its benefits without taking on more than is necessary.

V. CONCLUSION

In a competitive industry such as retail, big data is becoming ever more important, and failing to take advantage of big data can cause a business to be out-competed in the marketplace. Obtaining and analysing customer trends and many other business factors has been shown to provide a huge advantage in the form of data-driven decision making, as well as helping to reveal otherwise unknown business opportunities.

While this sounds too good to be true, there are certain challenges involved with adopting big data. As data becomes cheaper and easier to obtain, and big data is a rapidly growing industry, there is an alarming shortage of available data scientists which is predicted to grow towards the end of the decade. As recruiting data-scientists who do not have the required skills/experience/mind-set can be damaging to a business, it is vital to ensure that the correct approach is taken when planning for the future.

Firstly, it is essential to ensure that data-scientists are flexible and have a good “domain knowledge” in order to
understand business needs and provide answers to the right questions.

Secondly, it is important to integrate data scientists into the rest of the business so that their skills and experience are passed on to other employees who themselves may turn out to be a valuable future asset.

Thirdly, it is also sensible to gauge the required size of big data operations, as it is expensive and slow to analyse large quantities of data, and it is much more efficient if the same questions can be answered by analysing smaller quantities of data.

VI. REFERENCES


The Challenges Of Big Data

Mohammed Haroon
University of Derby
Information Technology
Derby, England
M.Haroon1@unimail.derby.ac.uk

Abstract— this is paper is going to discuss about how big data challenges are affecting small business enterprises. Businesses exploit big data without regard for issues such as legality, data quality, data meaning and process quality. This often can lead to poor decisions bearing great risks. Other concerns that are in question are privacy of data and scale. In this electronic age, increasing number of organizations face the problem of explosion of data. Small medium enterprises have looked to adopt big data within the enterprises but do they have the resources and employees that can make use of the data in order for the data to be of any value to them. The V’s which will be discussed in this paper will be Volume, Velocity, Variety and Value. With big data becoming a new attraction within the industry is it worth small medium enterprises trying to use big data to help their businesses.

Index Terms—Big Data, Security, SMEs, Frameworks, Challenges.

I. INTRODUCTION

Approximately 90% of digital data that we encounter today did not exist two years ago and is predicted to grow from 2.7 Zettabytes to 35 Zettabytes by the year 2020, (Malik 2013) big data is the next big step for the business world. “Big Data” refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse, (Manyika et.al 2011). Big data has emerged as we live in a society that makes increasing use of data intensive technology, (Patel, Birla & Nair 2012). Oracle (2013) define big data has three different types of data which are traditional enterprise data, machine generated/sensor data and social data. It is estimated that data volume is growing by 44 % per year and will carry on till at least 2020. There are various different sectors that have used big data to help add value to their businesses. Saniuta, Roman & Pop (2013) have stated that 64% of retailers would use big data to get a better understanding of consumers, 57% would use it have better targeted market campaigns and 53% would use big data to have better planning and decision making. The public sector have used big data to obtain a better understanding of customers and to have better planning and decision making. Other than small medium enterprises big data has attracted attention from the government and academia. Manyika et.al (2011) has stated that the potential value of big data to the health industry can is $300 billion which is double the total annual health care spending in Spain.

II. THE IMPORTANCE OF BIG DATA

When big data is analyzed in combination with the traditional data, enterprises can develop a more thorough and meaningful understanding of their business, which can create enhanced productivity, a stronger competitive position and a greater innovation all of which can have a significant impact on a business (Oracle 2013). Zhang (2013) also has stated that big data helps increase productivity in manufacturing and developing a competitive edge for business. He also explains that the objectives of big data are varied. They are more in conjunction with the objectives of big data stakeholders. Big data is becoming more and more important and many industries are cashing in to get the information they need in order for them to boost business. Retailers are a big example of big data users as they know they usually will know who will purchase their products. Now the use of social media and web log files from their ecommerce sites can help them understand that who did not buy their products and why they did not choose to. This has given them the ability to have more effective market campaigns and improve supply chain efficiency through precise demand planning. Social media sites such Twitter and Facebook would not exist without big data. Their business models require a personalized experience of web which can only be obtained by capturing and utilizing all available data about a user or member, (Oracle 2013). Not only are retailers cashing in but many other originations are monitoring the content of thousands of social media posts every day to uncover customer insight enabling them to profit from emerging trends and to deliver a rich customer experience, (Malik 2013). The importance of big data is also highlighted by Broersma (2014) who stated that the government has announced that they will invest £73 million in UK big data. This investment is for 55 projects related to data analytics and for the establishment of four research center’s at English universities. The UK government has said they are investing this because they will focus on big data as the driver of the British economy. The minister of university and science David Willets states that the “Making the most of large and complex data is a huge priority for government as it has the potential to transform public and private sector organizations” he also adds that “The new data research centers will help the UK grasp these opportunities and get ahead in the global race.”
III. CHALLENGES OF BIG DATA

Big data has its advantages for businesses and also it has its disadvantages. There are various different challenges that come along with big data. The various different issues that exist with big data are issues such as privacy and security issues, storage and processing issues, analytical challenges and technical challenges. Another issue that IT professionals face is the how they can filter the most important data from all the information that they have collected so that they can actually use that information to add value to the company, (Katal, Wazid & Goudar 2013). The main issues that will be put up for scrutiny is the privacy issues of big data and also look at the big data vs. privacy and security issues

A. Privacy And Security Issues

The biggest concern is the security and protection of sensitive information. Small medium enterprises are collecting, analysing and making decisions based analysis of huge amounts of data which have been obtained from various different sources and security in the process becoming increasingly important (Smith, 2013).

Enterprises are currently storing approximately 11–15 terabytes of security data a week. For enterprises, the ability to detect data breaches within minutes is critical in preventing data loss, yet only 35% of enterprises have said that they have the ability to do this (Info security, 2013). There have been cases where people’s data has been compromised such as UK based insurance and financial services firm Staysure had to notify more than 93,000 customers that their personal information which included encrypted payment card details was compromised (Roman 2014). Kaplan (2014) also had informed that customer’s payment information had been compromised at 34 Specs stores. More than half a million customers at 34 liquor stores owned by Specs may have had critical financial information stolen in a sophisticated computer scam that persisted for a year and a half which also shows that how important it is for companies to detect data breaches as fast as they can. Smith (2013) states that one of the largest data breaches in recent history involved Sony’s PlayStation Network in 2011. Experts estimate that the Sony’s related cost to the breach was between 2.7 and 24 billion dollars. According to BBC news (2014) Morrison’s staff had been affected due to payroll data being stolen by an employee. It is thought that details of 100,000 employees had been stolen the information included bank account details.

Advantech (2013) have said that the possible ways to enhancing security are by firstly having secure encryption technology which needs to be used to protect confidential data mainly for Personally Identifiable Information and Protected Health Information , but also a company’s own Intellectual Property. Secondly careful management of access to the crypography keys which unlock the encrypted data must be put in place.

B. Variety

The vast amount of data that is being produced which is not of a single category as it not only includes traditional data but also semi structured data from resources such as web pages, web log files and social media sites. All this data is unstructured which for small business enterprises is difficult to handle by the existing traditional analytic systems, (Katal, Wazid & Goudar 2013). (Demchenko et al 2013) also states that the variety tackles the issues of complexity of data and the information and semantic models behind the data. Kaisler et al (2012) suggests that from an analytical perspective, it is probably the biggest obstacle to effectively using large volumes of data. Incompatible data formats, non-aligned data structures and inconsistent data semantics represent significant challenges. Rarely does data present itself in a form perfectly ordered and ready for processing. The reality of data is that it’s messy. Different browsers send different data, users withhold information all this usually tends to lead to errors and inconsistency of data use (Dumbill, 2012).

C. Value

Data value measures how useful the data on hand is to help make decisions. (Kaisler et al, 2012). Users can run queries against the data that is stored and thus can conduct important results from the filtered data. These reports help to find business trends according to which they can strategies. The true value of data can only be assed if small enterprises have the relevant people to who can actually create value from this data. (Katal, Wazid & Goudar 2013). Zheng, Zhu & Lyu (2013) explains that big data will have a massive effect on big data employment where 140,000-190,000 workers with deep analytical experience will be needed furthermore 1.5 million managers will need to become data literate. Madkour, Aref and Basalamah (2013) have argued that the issue in learning on big data lies in understanding what the data means. This is known as semantic interpretation as it deals with the interpretation of ambiguous and imprecise data. To get the true value of big data it is important that the true interpretation of the data is made as then only will the data help make vital business decisions. With big data becoming an important aspect of business there is a limited amount of employees that can actually make use of this data. Maik (2013) has suggested that increasing volume and speed of data generation can result in decreasing data certainty. The issues of data quality must be addressed because of the incomplete and uncertain nature of big data.

D. Volume

Big data volume includes such features as size, scale and amount, (Demchenko et al 2013). Dong & Srivastava (2013) have said that not only can each data source contain a huge volume of data but also the number of data sources even for a single domain had grown to be in the tens of thousands which shows how much volume of data that now exists. Kaisler et al (2012) argue that data volume measures the amount of data available to an enterprise which does not own that data as long as they can access it, however as data volumes are growing the value of different data records will decrease in proportion to age type richness and quantity amongst other factors.
E. Velocity

Velocity is the speed of how fast data flows from different sources like business process machines, social media sites and mobile phones. The flow of the data is huge and more importantly continuous. This real data can help researchers and more importantly enterprises make valuable decisions that provide strategic competitive advantages and return on investment of you are able to handle the velocity (Normandieu, 2013). Dumbill (2012) has stated that the increasing rates at which data flows to businesses followed a similar pattern to that of volume. The problem was that the big organisations had the system that were able to cope with fast moving data to their advantage but small medium enterprises need to have that in order for them to use big data to their advantage.

In the next section ways of how issues can be addressed using the relevant frameworks will be discussed. Frameworks that can be used will be specified and how they will help small medium enterprises tackle the issues that they may have. The main challenges that can arise for small medium enterprises are such things as data being stolen and not knowing the true value of data. Using the framework a discussion will also be on what steps a business can take for them to overcome these issues and if certain issues keep arising what they can do and have in place for them to recover as fast as possible which as a result could mean less people could be affected.

IV. ISO 27002:2013

This international standard is designed for organisations to use as a reference for selecting controls within the process of implementing an information security management system (ISMS). This standard is also intended for use in developing industry and organisation specific information security management guidelines taking into consideration their specific information and security risk environment.

Big data’s highest risk is that important and sensitive data does not get stolen or lost. For small medium enterprises it is important because this information is what they have to use in order for them to gain competitive advantage. (ISO 2013) Point 11.1.5 is a method of how small medium enterprises can avoid any data from being taken. This could be avoided as it states that there should not be any unsupervised pupils working in areas where data could be taken. It also states that vacant areas should be physically locked. Other issues that may arise with small medium enterprises is that they may have old machines which are not able to analyse the data fast enough or they may not be able to process the data fast enough 11.2.4 has explained that equipment should be maintained in accordance with the supplier recommended service intervals. This shall allow the system to perform at the highest level on a constant basis due to regular services that will take place.

Another issue was that about how making sure that all the places where the information is kept is safe. Staff members who have log in credentials need to make sure that they have secure environments when they need to log in so that their information does not get compromised. Small medium enterprises also need to keep all the information away from unauthorized individuals. (ISO, 2013) 12.4.2 guideline says that logging facilities and log information should be protected against tampering and unauthorized access. (ISO, 2013) 13.2.4 states that confidentiality or non-disclosure agreements should address the requirements to protect confidential information using legally enforceable terms. This means that enterprises need to state all the important information such as staff cannot share the credentials to unauthorized individuals and that all the data that they have access needs to stay private and for work purposes only.

V. ISO 22301

A successful business has to make sure that it has a plan for the expected and the unexpected. Disruptions to any business can result in data risk, revenue loss and failure to deliver a service at all. This is why a business should have strong business continuity planning. ISO 22301 specifies the requirements for setting up and managing an effective business continuity management system (BCMS) for any organization, regardless of type or size. ISO 22301 states that every organisation should have system in place to avoid excessive downtime and reduced productivity in the event of an interruption (Sharp, 2012). ISO also helps to identify potential threats that a business may have and the impact these threats may have to business operations. In regards to big data we gave the issues that could arise for enterprises. Issues such as data getting compromised can occur for an organisation for example such as what happened at Morrison’s this could result in productivity issues, ISO 22301 has a clause 7.3 which is awareness (ISO 2012). This states that persons doing work under the organisations control should be aware of the business continuity policy, their contributions to the effectiveness of the BCMS including the benefits of improved business continuity, the implications of not conforming with the BCMS requirements and their own role during disruptive incidents. What this means is that when issues such as this occur the staff members know what they have to do in order for the business to continue trading. (ISO 2012). (ISO, 2012) 8.4.4 Business continuity plans is about how the business shall establish documented processes to take when an incident occurs this will state how it will continue or recover from the incident within the allocated timeframes. The plans will contain what roles and responsibilities staff have during and following the incident, a process for activating the response and details to manage the immediate consequences of the incident in regard to the welfare of individuals who may have been affected, tactical and operational options for responding to the disruptions and prevention of further loss. This is vital for a business because if there are guidelines that a business has already noted of what steps the business takes to tackle an issue this will then help the business overcome this issue as fast as possible which as a result could mean that less people may have been affected by the incident. When a plan is made it’s important for a business to keep evaluating that plan which is stated in 9.1.2 that a business evaluates the procedures in order for the business to ensure their continuing suitability and adequacy and effectiveness. this will be done by the management of the business as they will audit the results they
have found and take measures from then to see if the BCMS do need improving (ISO 2012).

VI. CONCLUSION

This paper has discussed about some of the challenges that small medium enterprises face when looking to use big data. The big organisations have used big data very effectively and are receiving the rewards. Big data is a valuable asset to use as it can help the business in many aspects such as sales and also productivity but there are issues that need to be overcome. Small medium enterprises do not have the financial power of big organisation, so will they be able to afford the system that can help use big data to their advantage. Another issue that we have is, are there enough people who can actually understand the data that is put front of them. Having the relevant system in place is good but then having someone there who cannot determine the true value of the data will be of no relevance to the business, it will result in incorrect business decisions which could potentially result in big financial loses which small medium enterprises simply cannot afford. Big data is a tool which small business should use but they need to make sure that they can get the benefits that they need. Important aspects that small medium enterprise need to consider are making sure that all the data they have is secure and it cannot be compromised and they have the personnel who can understand the data therefore making the data of value to the company and receiving the benefits of using big data.

VII. REFERENCES


Overcoming The Talent Shortage In Big Data Analytics
A Guide For SMEs

Robert Heeley
University of Derby
Derby, United Kingdom
R.Heeley1@unimail.derby.ac.uk

Abstract—Time and cost-efficient analytics of Big Data has become a key driver for success in the competitive business world. As the volume of generated data increases, so do the number of companies looking to leverage value by adding big data initiatives to their business. The demand for talented data scientists is outstripping the supply, and recent reports have shown that this situation is only going to get worse. Companies have it within their power to circumvent the talent bottleneck through forward thinking and a willingness to adapt. The solutions offered in this report are by no means exhaustive, but they provide a good foundation for those willing to build upon it.

Index Terms—Big Data, Data Science, Data Scientist, Analytics

I. INTRODUCTION

In the competitive business world, time and cost-efficient analytics of Big Data has become a key driver for success (Cohen et al. 2009); with both established competitors and new starters alike leveraging data-driven strategies to innovate, compete, and capture value in a bid to outperform their peers (McGuire et al. 2012).

The sheer volume of data being generated annually is increasing at a staggering rate; with experts predicting that by 2020, the digital universe will total in at 35 zettabytes; 44 times larger than in 2009 (IDC, 2011). Though this boom has the potential to grant excellent insights for the companies embracing the Big Data phenomenon, there is a caveat; a lack of skilled workers that can make sense of the large quantities of data (Staffing Industry Analysts, 2013).

Companies wishing to leverage maximum value from their Big Data need to tackle the issue head on, putting the recruitment and/or training of skilled data scientists at the forefront of their Big Data plans.

II. THE LACK OF BIG DATA TALENT

It is unlikely that anybody with a keen interest in Big Data analytics is ignorant of the shortage of talented workers required to fulfil all of the related job roles as more and more companies are tuning into the benefits of Big Data analytics.

Research carried out by the McKinsey Global Institute in 2011 forecasts that by 2018, the United States alone could be facing a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million managers and analysts with the knowledge to make effective decisions from the analysis of Big Data (Manyika et al. 2011). A report released more recently by e-skills UK, and SAS UK, a leader in business analytics, predicts that the demand for big data specialists in the UK alone is set to increase by 243 per cent over a five year period to 69,000 (e-skills UK, 2013). The report also recognises the impact that the rise of big data has had beyond that of big data specialists, to the users of big data. These are the employees outside of the IT or data teams that make use of the big data via tools such as dashboards, key performance indicator data, and market analysis. The number of big data users is predicted to increase by 177 per cent over the same five year period to a total of around 644,000 by 2017. So what has led to this widening hole in the employment market? There are no doubt a whole slew of reasons that amalgamate to form the bigger issue, certainly too many to include them in a single section of this report. Rather, this report will focus on the most widely recognised amongst the causes.

A. Increasing Uptake

Big data analytics is still a relatively new concept to the business world, but one that is rapidly gaining steam due to the highly publicised successes of big data initiatives (CSC, 2014). As a result, it is not only the new companies being formed around big data analytics that are looking for staff, but also existing companies that are looking to leverage value by adding big data initiatives to their business. The aforementioned report by e-skills UK and SAS UK (2013) predicts that around a third of the UK’s larger organisations will be implementing big data analytics programs by 2017, equating to roughly 6,400 organisations that will be on the hunt for talented employees, and that is just the larger organisations (employing 100 or more staff) and does not include SMEs and smaller companies. Taking those numbers for the UK alone, it is easy to imagine just how large the global deficit will become.

B. A Lack Of Education

In 2012 the Wall Street Journal published an article regarding the shortage of talent related to Big Data analytics. Amongst the sources for that article was Nigel Shadbolt, the co-director of the UK’s Open Data Institute and professor of artificial intelligence at the University of Southampton.
Shadbolt observed that courses for data science did not yet exist; though some topics relating to data science may be covered across various departments around the country, data science as an integrated discipline was only just starting to emerge (WSJ, 2012). Since that article, a number of universities have added Data Science related degrees to their prospectus. The vast majority of these are postgraduate degrees looking to build upon a previous degree in computer science, engineering, physics or mathematics, although other disciplines such as psychology, economics, and health may be considered, provided that the student possesses demonstrable mathematical aptitude (City University London, 2014). There are also undergraduate degrees emerging such as the BSc Data Science at the University of Warwick, organised as a joint venture by the Department of Statistics and the Department of Computer Science (Warwick University, 2014). Similarly to the postgraduate degrees, mathematical skills are a staple of the course. With these courses now in place and more sure to follow, the flow of data scientists should see an increase. However, with the data science courses being new and unlikely to garner the popularity of more generalised degrees such as Computer Science, this increase is unlikely to stem the tide of data scientist vacancies.

C. The Data Scientist Assignation Is A New One

When a company takes steps to implement a big data initiative, they may be tempted to launch into recruitment for “Data Scientists”; after all, these are the talented individuals with all of the skills to make the initiative a success. Though this is actually true to some extent, the issue lies in the job title. The title of “Data Scientist” was allegedly coined by Jeff Hammerbacher back in 2008 to describe the roles and skillsets of the team that he led at social networking giant Facebook (Simon, 2013). The merging of the research scientist and data analytics roles led to the new assignation that has been adopted by the rest of the Big Data community. A look at the popular job search engine indeed.com demonstrates the birth of the data scientist role.

The title is now something of a staple and many employers would be glad to call themselves a data scientist; they are in demand, and for an employee, being in demand can only lead to opportunities. However there may well be talented individuals flying under the radar in the guise of “Research Scientist” or “Data Analyst” that are equally qualified and capable of getting the project on track.

III. HOW TO OVERCOME THE TALENT SHORTAGE

Data scientists are the experts responsible for designing the intricate models, algorithms and visualisations that enable their companies to distil insights from the huge mass of available data. However, crunching data is not the sole purpose of a data scientist; companies are also looking for individuals that are able to understand the “Business Challenge”, create valuable and actionable insights to the data, and effectively communicate their findings to the business (Harris et al. 2014). The widening gap between the number of data science roles and skilled workers to fill them is well documented, and if current predictions are anything to go by then the situation is only likely to get worse (e-skills UK, 2013). Although there is no magic wand to conjure up skilled workers for those companies looking to recruit, there have been some suggestions publicised that may help to alleviate the issue for the companies willing to take the necessary steps.

A. Training Up Fresh Graduates

Data science courses may still be few and far between, but the flow of smart individuals graduating in other courses from universities across the globe is unhindered. With mathematical skills being one of the most requested skills amongst data scientists, a company can work with university career teams to hire smart numerate people from multiple disciplines that can then be trained up internally to round off their skill sets. This is the approach that UK supermarket giant Tesco utilises, and claims that they have no trouble finding the skilled IT professionals that they require (V3, 2013). This is no mean boast, with Tesco being somewhat of a pioneer of big data across the UK (Forbes, 2013); projects ranging from targeted marketing through their Clubcard data, to an estimated cut in cooling costs by up to 20 million euros by studying refrigeration data (Goodwin, 2013).

B. Create A Data Science Team

One of, if not the biggest reason for the lack of talented data scientists lay in the scarce combination of skills that are required for the job. A data scientist is expected to have mastered advanced statistical and quantitative methods and tools, computing environments, and the languages and techniques required for managing and integrating large data sets. Further to these technical skills, data scientists must also possess intricate knowledge of the industry as well as the business acumen to create appropriate models, and excellent communication and data visualisation skills in order to explain their models and findings to others.

An option available to those companies struggling to find the rare individuals that possess all of these skills is to recruit a team of varied professionals that will collectively tick all of the required boxes (Accenture, 2013). This solution not only provides the expertise required to leverage value from Big Data, but also shares the burgeoning workload between team members. Having members of varying expertise collaborating on a single goal can also provide fresh insights that may be out
of perspective of any one individual. This observation has led to the suggestion that recruiters may wish to look beyond their industry and consider recruiting from other disciplines such as physics majors for their mathematical imagination, even artists such as graphics designers that can bring a new level of creativity and imagination to data visualisation (Accenture, 2013). The image below highlights the data science strengths of five separate disciplines that come together to fill the data scientist role.

Fig. 2. Roles in Big Data Analytics

C. Look Beyond Domain-Specific Languages

Software engineering for the purposes of data manipulation, computational statistics, and visualisation, has for the past decade or so largely fallen into the corner of the domain-specific programming languages, with “R” gaining the lion’s share. R is an open source programming language with huge community support that has led to over 6000 packages being created to support virtually every data manipulation, statistical model, and chart that the modern data scientist could ever need; or so claim Revolution Analytics (2014).

When it comes to the realm of big data however things are not quite so clear cut. The R programming language relies on system memory to handle data sets and objects, and since there is a physical limitation to a computer’s memory, it falls to reason that there is a limitation in the size of data sets R can manipulate. Any attempt to load a bigger data set into R will lead to an inevitable system crash (McKelvey, 2013).

Another issue with using a domain-specific language such as R is in the complexity of learning the language as highlighted by Bob Muenchen (2012). Part of the reason people struggle to learn R is that it does not resemble traditional programming languages. As R expert John Cook describes it:

“It is an interactive environment for doing statistics. I find it more helpful to think of R as having a programming language than being a programming language” – (Cook, 2014)

When companies are already struggling to fill their ranks with capable data scientists, limiting scope to a niche community of domain-specific developers would be like swimming against the tide.

D. Embrace The Shift To Python

Python is a powerful open-source programming language that is renowned for its flexibility and simplicity; being useful for general-purpose programming as well as analytical and quantitative computing (Continuum Analytics. 2014). Python is one of the most popular programming languages in the world (TIOBE Software. 2014), and garners praise for facilitating productivity, quality, and maintainability of software (Bird et al. 2009).

Python has long been great for data wrangling and preparation, but less so for data analysis and modelling. This is no longer the case, with tools and libraries being developed to allow users to carry out their entire data analysis workflow in Python rather than switch to a more domain specific language such as “R”.

The advent of the “iPython” interactive computing environment by Perez (2007) has allowed scientists worldwide to run their experiments and get results in real-time; granting the ability to visualise the data in a range of ways (Ravven 2013). This project is now refined and expanded upon by hundreds of contributors and is used by scientists of all disciplines around the world, and together with libraries such as “SciPy” and “Pandas” provides much of the mathematical functionality typically found in the “R” programming language.

Engineers from the real-time online advertising-platform company “AppNexus” explained the importance of the Python programming language in the scaling of their operations (Jackson, 2012). David Himrod, the Director of Optimisation and Analytics for AppNexus stated that the key to Python’s usefulness was in its simplicity. The easy to learn nature of the language had enabled employees of differing backgrounds such as engineers, mathematicians and analysts, to work together on the same technology stack, prototyping new functionality for the company. Himrod further highlighted the intrinsic benefit of no longer having to recruit engineers from specific programming backgrounds, as Python can be easily taught to new employees.

IV. Conclusion

The talent shortage for big data analytics is a major obstacle for any company looking to break into the world of big data, and judging by the predictions from McKinsey and e-skills UK the problem is going to get worse before it gets better.
Although a company cannot avoid the issue completely, that is not to say that it cannot be navigated around with some careful planning.

By utilising a combination of the proposed solutions in this report, a company should not only be able to minimise the impact of the talent shortage on their big data plans, but even turn it into their favour.

By recruiting promising graduates straight from the universities, a company can gain ambitious, knowledge hungry talent that can learn the intricacies of the business as they grow; sound business knowledge being a highly desirable trait in a data scientist. Graduates are also far more affordable than established data scientists, which in the current economic climate can demand a premium for their services.

Possible the biggest boon to a company looking to take on larger projects is the ability to put together a data science team. By broadening the scope of recruitment to multiple disciplines, not only can a company cease the struggle to find individuals in possession of all desired data science skills, but this also carries the added benefit of bringing together a wider array of knowledge and expertise that may produce even more insights from the big data.

Whether a company utilises a team of data scientists or only one or two individuals; utilisation of the Python programming language can make it relatively easy to get employees up to speed and working on the same technology stack as other departments in the company. This can drastically minimise the development time of new software, and allow individuals to implement their own algorithms and models without the need of a specialist developer.

V. REFERENCES


How Green Is Big Data Computing?
What Can Be Done To Reduce Its Environmental Impact?

J. Hill
University of Derby
Derby, England
j.hill6@unimail.derby.ac.uk

Abstract—This paper explores the use of data centres and the potential environmental impact they have, techniques used by large innovative companies to increase efficiency of energy used within data centres and additional strategies these larger companies employ to further limit their carbon footprint on the planet by using data centres.

Index Terms – green computing, data centre, efficiency, energy usage, environmental impact.

I. INTRODUCTION

Big data computing these days are used in many different places, from web applications like Facebook and Google+ to e-commerce at a huge scale such as Amazon and ebay. Each leveraging the advantages to be gained from data collected from their customers. Due to the nature of the big data computing a company would require the storage of vast quantities of data and huge amounts of processing power to sift through it and analyze it. The solution to this is the use of a data centre essentially a distributed system of many servers housed in a controlled environment.

This article will analyse and explore one specific issue that surrounds the use of big data computing, this of course is as the title suggests that of an environmental one. These data centres are usually very large in size; contain hundreds if not thousands of computers and use large volumes of electricity, what impact does this have on the environment? How efficient is the use of this electricity? What could be done to increase energy efficiency used at these data centres? These are all questions I am hoping to answer in my article.

II. THE ENERGY IMPACT

It is a safe assumption to make that in today’s age data centres are somewhat of a necessity. They are frequently used to underpin a multitude of systems including large scale logistic operations, high intensive data storage of web applications, big data analysis for targeted advertisement campaigns and IT cloud based office applications and services. With the advantages that these systems bring at what price are we paying for it in terms of environmental impact?

In 2007 56 TWh of power had been consumed by data centres in Western Europe alone, by 2020 this is estimated to be almost doubled. The US tells a similar story, in 2006 data centres made up for 1.5% of all power consumed in the U.S. (Bouley 2010). This percentage has climbed every year since and is now closer to 2% of all power used worldwide (Gregory 2013).

III. THE ENVIRONMENTAL IMPACT

So with all the massive amounts of energy consumption what is the actual impact on the environment? According to a paper on data centre efficiency in the U.S. their data centre consumption contributes 76 million metric tons of CO₂e adding to the greenhouse gases impact annually (Costello et al. 2012).

The footprint is further increased with where the data centre acquires its electricity; the issue here is that data centres may typically be fully reliant on coal, uranium or natural gas power plants due to their ability to react quickly to increases and decreases in demand. Due to the fact that Coal supplies 40% of global electricity it’s no surprise data centres typically draw supply from coal power stations. In fact some of the largest data centre companies such as Apple, HP, IBM and Oracle share over 50% of their electricity from coal (Mills 2013). Whilst this may not be helped situating a data centre in an area reliant on coal does encourage its use and some companies such as Google attempt to locate data centres near renewable energy sources.

Additionally on top of this effect the resources used to construct data centres is astonishing, taking a look at a typical 1MW data centre over a 10 year lifespan (not including the building) materials used are as follows:

- Electricity 177,000,000 kWh
- Water 227,000,000 l
- Copper 65,771 kg
- Lead 9,525 kg
- Plastic 14,968 kg
- Solder 5,443 kg
- Steel 171,004 kg

IV. PERSPECTIVE

To gain a perspective over what these numbers all mean comparing it to an office building appears quite useful. A data centre requires effective cooling of large amounts of generated heat whilst powering networking equipment and 1000’s of servers. A data centre has typically a 10 year life cycle whereas offices 40 years. Also a data centre uses 40 times more energy as an office building of similar size. Being the monstrosities they are data centres could be better described as industrial factories rather than office buildings.
V. DATA CENTRE EFFICIENCY

There are many variables that affect a data centres efficiency, some of the important ones are location, IT equipment/load and electrical efficiency of supporting hardware.

A. Location

The climate a data centre is in has a huge impact on its efficiency, the reason for this is that for computing equipment to run optimally it must remain at a cooler temperature. So in a hot climate chilling server rooms or racks becomes a bigger burden and uses more energy. In a colder climate not only is it easier on cooling requirements but cold air from outside may be used instead of big chilling units.

In one of Goggle’s data centres in Hamina they make use of cool sea water pumped into heat exchangers in turn removing heat from warmed water direct from servers. This enables the site to use no refrigeration which is the traditional less energy efficient approach.

This is one of the problems data centre engineers face is that through air conditioning techniques it is all too easy to overcool the entire room due to warranties of hardware which has implications on energy requirements (Patterson et al 2009).

B. I.T. Equipment/Load

This variable is focused on the hardware itself. One of the biggest aspects of this is server underutilisation, due to often short bursts of activity SLA’s (Service Level Agreements) over compensate for the maximum foreseen load. Thus in turn this leads to the underutilisation alluded to earlier. The problem of efficiency isn’t just purchasing too many servers but of keeping idle servers running. Idle servers still draw 60% of peak power, wasting large amounts of energy and cooling efforts (Meisner 2009).

In addition I.T. equipment itself may be inefficient typical server PSU’s can be found to waste 25% just converting current to DC (Google 2009) The same can be said for even the most efficient UPS’s wasting 5-8% of energy by using centralised UPS rooms (Shankland 2009).

C. Electrical Efficiency

Electrical efficiency is concerned with supporting equipment of the data centre and how much of the electricity used is actually consumed by servers etc. So to understand electrical efficiency how much of the power is used for cooling, lighting, UPS, generators and supporting facilities compared with power used on servers, storage and networking equipment (Uddin 2013). The industry standard measurement for this is called power usage effectiveness (PUE), which I will cover in the proceeding section.

VI. MEASUREMENTS

To measure a data centre’s efficiency a number of metrics have been created as industry standards, this enables the ability to use the metrics as a benchmark to work towards to improve efficiency.

- PUE – This metric is widely used by many organisations with data centres, PUE measures power usage effectiveness. As an example a PUE of 2.0 means for every watt of power an additional watt is used on cooling, UPS, Lighting etc (Uddin 2013). According to the Uptime Institutes research in 2012 the average PUE was 1.89 and just fewer than 10% of respondents reported 2.5 PUE or greater (Uptime Institute 2012). This suggests a lot more can be done to create efficiency in a practice that already consumes 2% of worldwide power usage.
  - CUE – Carbon usage effectiveness is the metric used in combination with PUE. Essentially it measures total CO² emissions caused by total data centre energy divided by IT equipment energy. This enables a data centre to determine how much carbon they use for the PUE they have, in simplicity this metric doesn’t just measure energy efficiency but how much CO² is produced via energy sources and other equipment on the site (Belady 2010).

This by no means is the definitive list of data centre metrics but makes a decent start in understanding efficiency as these incorporate aspects of the entire site while other metrics might focus on specifics like cooling effectiveness etc.

VII. WHAT CAN BE EMPLOYED TO INCREASE EFFICIENCY

There are a number of techniques that may be used to increase a data centre’s efficiency:

A. Cold-aisle Containment

This is a practice used to in short stop cold air used for cooling mixing with warm air given off by servers. By creating a barrier to the front end of the aisle and the roof between racks it contains what is known as the cold aisle created by under floor cool air fed by the chillers. Servers draw cold air in front and vent heat out the back into the hot aisle; this then naturally flows back to the chillers. Because of this containment or controlled air flow it enables the tuning of the chillers to raise temperatures up to 27ºC due to cold air being more efficiently used. This in turn is proven to reduce energy costs by 10-20% for a traditional uncontained air flow data centre (Daniels 2013).

This efficient use of cooling is especially important due to rack densities increasing with the use of blade servers which enables data centres to fit many more servers per rack but in turn increasing heat produced (Murrill 2011)
B. UPS and Power Distribution Efficiency

In a traditional setup data centres contain large UPS rooms to support servers during a power outage. UPS’s and servers use DC current which in the traditional setup AC is converted to DC in two stages, once for UPS and again for servers via the PSU. It is known that PSU’s have efficiency between 60%-90% dependant on load and manufacturer. UPS’s have efficiency in power conversion on average between 86%-90% dependant on load, manufacturer and setup (Meisner 2009; Ton 2009).

A solution to tackle this loss of energy in conversion is to decentralise UPS rooms and install batteries on each server, this not only removes one stage of relatively inefficient conversion but also cuts down on UPS power suffering from inefficiencies due to the length power travels from UPS rooms to servers. Google implemented this solution and estimated annual savings of 500kWH per server (Google 2009).

C. Consolidation via Virtualisation

According to research the average data centre only utilises 20-30% of servers (Meisner 2009). In further research by VMWare some servers are found only to utilise 0.5% of its processor (Talaber 2009). The strategy of consolidation entails putting more load on each server and switching idle servers off. Due to idle servers still drawing 60% of power there is quite a considerable saving to be made (Meisner 2009). Criticisms of this strategy may suggest that a server consumes more power, this is a true statement but on a machine that is at 10% capacity an additional 10% load only increases power consumed by 15 watts. A far cry from having two of the same servers running at 10% costing the combined total of 320 watts (Talaber 2009).

![Fig. 2. CPU Utilization and Power Usage](Image)

However shipping workload onto other servers isn’t as straightforward, this is because each server may only be utilised for one purpose so this is where virtualisation comes in. Virtualisation enables the ability to host a small number of operating systems on a single server. Thus in turn on server can host a multitude of services that may have been originally split across many servers. This then enables underutilised servers to be turned off and increases a server’s utilisation.

D. Water Cooling

This is another strategy that is employed by some data centres to further increase energy efficiency. The nature of water cooling is to run water passed warm components to take the heat away. Water cooling is said to be 4000 times more efficient than air cooling (Chapman 2012). In some cases warm water is even fed into the heating system of office areas to remove the need for a boiler for heating again saving energy. In some cases cool water is chilled either by cooling ponds/towers or fed by sea water, this removes having to chill water to be fed back into the system (Google 2009).

E. Shared Storage

Sharing disk space across a cluster or more servers can significantly save energy. This moves having two or more underutilised disks per server. Using a NAS or a SAN allows significantly save energy. This moves having two or more underutilised servers to host a multitude of services that may have been originally split operating systems on a single server. Thus on server can host a multitude of services that may have been originally split across many servers. This then enables underutilised servers to be turned off and increases a server’s utilisation.

E. Shared Storage

Sharing disk space across a cluster or more servers can significantly save energy. This moves having two or more underutilised disks per server. Using a NAS or a SAN allows significantly save energy. This moves having two or more underutilised servers to host a multitude of services that may have been originally split operating systems on a single server. Thus on server can host a multitude of services that may have been originally split across many servers. This then enables underutilised servers to be turned off and increases a server’s utilisation.

F. Intelligent Management

When data centres are designed there are many considerations to take into account. Usually decisions are influenced by power grid stability and Internet connectivity. Some companies vowing to go green however are also influenced by local renewable energy sources, energy efficient cooling from the climate and possibilities for on-site renewable energy generation.

Facebook’s data centre in Lulea takes advantage of local hydro plants to completely power their 30,000 sq m data centre. Additionally it uses air from outside to cool the facility 10 months of the year (Gregory 2013).

In another example Apple’s data centre in maiden North Carolina made use of its location by creating a 100 acre solar farm which powers the data centre by 60% (Burrows 2013).

Given data centres reputation for energy consumption steps taken like situting them next to renewable energy sources they are certainly paving the way for others and investing in companies associated with renewable energy production.

IX. Conclusion

In conclusion to this research paper managing these areas in which have large energy inefficiencies not only tackles the impact data centres have on the environment but also enables cost savings. Overcooling and the use of greener energy providers would make a formidable start in reduction of environmental impact. There are likely scenarios in smaller companies that some of these energy saving techniques are not possible. Research suggests in the U.S. energy wasted in smaller server rooms and cupboards onsite due to poor operation strategy is responsible for $2 billion dollars electricity costs (Natural Resource Defense Council 2012). This is when companies like this may consider outsourcing...
services to cloud providers not only for an environmentally friendly solution but cost too. Research by Google suggests a company that migrates to a cloud solution or provider could save 68%-87% in energy and reduce similar quantities in carbon emissions (Google 2012).

Given these gains for both the environmentalist and data centre companies there is little reason that a data centre provider or smaller company cannot improve efficiency and reduce environmental impact that inevitably affects World energy consumption increases year on year.

X. REFERENCES


Big Data Analysis: Is Web Analytics Sufficient For SME’s?

James Adam Hunt
School of Computing and Mathematics
University of Derby
Derby, Derbyshire
j.hunt6@unimail.derby.ac.uk

Abstract— With 20% of UK business admitting they have poor or very poor understanding of the issues surrounding Big Data, this report argues that web analytics is a suitable Big Data solution for Small to Medium Enterprises (SME’s). It does this through explaining both concepts and comparing their advantages and disadvantages. Finally, a conclusion is drawn with the facts raised in these sections to aid a manager or director of an SME in choosing an apt solution for their market analysis.

Index Terms—Web Analytics, Big Data, Comparison

I. INTRODUCTION

The world has become more connected through the ever expanding internet and each day exabytes of data is generated. In fact, it is predicted that by 2015 over 83.8 exabytes of data will be transmitted through the internet. (Cisco, 2013) This data can show what is popular, what is important and, through analysis, future trends. This makes acquiring such an advantage very desirable for companies as insights into current and future trends can help improve profits. This can be seen by the fact that 97% of companies with revenues exceeding $100 million use some form of business analytics. (Chen and Chiang et al., 2012)

As the data being produced is in vast quantities, finding the trends you are looking for can be difficult, especially for Small to Medium Enterprises (SME’s) with limited IT resources. Due to this, finding the right analytic solution is a very important task. There are many approaches a company has to consider such as whether to build a network in-house to process the data for their needs, or outsourcing to an independent company. Not all data analytic solutions require massive amounts of processing power or highly trained teams to understand. 20% of UK businesses admitted that their understanding of Big Data falls under the categories of either poor or very poor. (SAS, 2014, p. 8) A solution that offers deep, useful insight into a company with much less difficulty would therefore be more helpful, one such solution is web traffic analysis.

Rather than focusing on the public as a whole, web traffic analysis focuses on a company’s current consumer base. Figuring out where they come from and how they found your company to begin with can be a helpful insight into discovering new customers and keeping them. (Phippen and Sheppard et al., 2004, p. 284-293)

II. BACKGROUND

When it comes to pursuing the concept of Big Data to improve your SME, there are many different approaches that have different costs and training requirements. When deciding for your business, these variables need to be considered.

A. Web Analytics

Web Analytics is used throughout the internet. In fact in 2009, over 48% of all websites used Google Analytics (W3techs.com, 2014), not including the many other companies who offer analytics services. Web analytics works through augmenting a company’s website to record visitor data. A small script is placed on each page which records visitor data and sends it to Google for processing. (Google Analytics, n.d.) This is then correlated with other visitor data to develop trends, such as how many people visited at a certain time or which country contains the majority of your visitors. This is combined with Google’s own information to discover what search terms led them to the website.

B. Other Forms of Big Data

As there is a constant amount of data streaming in from a large number of sources, large companies such as IBM offer to process it for you. (IBM, n.d.) This Analytics-As-A-Service allows an SME to have the business insights of a much larger corporation for a fraction of the cost. Big Data can involve more than just customer information and trends, it can also involve how your business runs and performs. UPS utilised telematics sensors in more than 46,000 vehicles which were analysed to calculate their routes. This was then used to refine them, cutting 85 million miles off of daily routes and saving 8.4 million gallons of fuel. (Davenport and Dyché, 2013, p4)

C. What This Article Entails And Why It Is Important

No one Big Data solution can be claimed better than the rest, but each has a different purpose. For example, an SME with a large web presence could find all the insight they need in web analytics rather than having to branch out to either acquire their own network for the purposes of Big Data analysis or hire an outsourced company to do it for them. Some forms of Big Data can be combined to improve the outcome received from processing them, such as combining a months’ worth of visitor data with a months’ worth of e-Commerce sales data to
As the general populace is connecting to the internet more and more, with over 2.7 billion in 2013 (ITU, 2013), a web presence is an important step for an SME and being able to analyse how people find you could be the most important resource of all. So although no one form of Big Data can be claimed better than the rest, web analytics could be a staple for web-based SME’s through its availability, cost and usefulness.

III. INVESTIGATION

As outlined above, different forms of Big Data analysis are used for different purposes, whether gathering customers, improving relations or cost reduction within a business. Due to this, there are many approaches to Big Data, including the sources of data used and how they are analysed, each having different costs and results. The purpose of this section is to analyse a few of these types as well as their effectiveness and efficiency compared to web analytics alone.

A. Current Use Of Web Analysis

The data obtained from web analytics can be used in many ways and suppliers of these analytics tools are continuously adding new features to make it more useful. Commonly, these tools consist of the ability to view anonymous user data such as approximate location, time of access and what page was accessed. (Google Analytics, n.d.) Some tools expand on this, such as Google Analytics which branches out into multiple devices and multiple platforms. Custom tools can also be created using an in-built API, allowing the data to be processed in any way needed.

The benefit of this approach is it is focused to the SME. The trends found aren’t globally applicable but instead apply solely to their business. This aids in transforming the idea of a positive trend into a positive change for the business as the trend you’re working towards has its roots in your business alone. For example, if the web analytics software found a majority of people found the businesses website through searching the keyword “Computer Repair” but were from another country, as a computer repair business you have a good hold of the search engine market, but to increase profits you would need to re-focus into the local area. Another example would be following visitors’ browsing.

Data is accumulated using cookies that are stored on the users system. This worked well up until 2012 when the public found cookies were being used by certain advertising companies to track their web usage as an individual, which led to cookies being erased by people often. (Information Commissioner’s Office, 2011) A study found these cookies were deleted by 31% of computer users every month which impacts negatively on the tools ability to trace visitors. (Abraham and Meierhoefer et al., 2007, p.2-3)

B. Outline Other Ways To Perform Big Data Analysis

As outlined above, there are many other forms of Big Data analysis, and they can be broken down into 3 components: where the data comes from, how the data is processed, and what is the end goal.

The source of the data can range from sales on your website to public tweets involving the most outlandish topics. To fit into the category of Big Data, the dataset need only be too much for a single machine to handle, but to suit a purpose you must find a dataset that corresponds to your end goal. (Russom and Others, 2011) For example, if an SME produces Ethernet cables and their end goal is to find if people have trouble with them, a good dataset to start from would be technology-based twitter feeds to find qualitative tweets that can be analysed into a useful end goal.

Processing the data is the next big step. As the size of the dataset is substantially more than a single computer can handle, the SME must choose between performing the analysis in house of using a cloud based service. If the in-house option is chosen, it will require training, equipment and extra staff, the cost of which may outweigh the benefit. The other option is a cloud-based solution, hiring a company to process and analyse the data for you using their own equipment. This approach would only require a trained marketing team to understand the results and liaison with the hired third party but comes with additional problems such as security of data. (Agrawal and Das et al., 2011, p. 530-533)

The end goal wanted is the key to the Big Data concept. Considering the amount of data available and the many ways of interpreting it, knowing what you’re looking for will save time. Time, in the world of processing power, is money, so the more efficient your search for the trend is the better.

IV. COMPARISON

To present the differences between web analytics and other forms of Big Data for SME’s, 3 key parts of each should be compared. These parts are its cost, the usefulness of the data produced and the difficulty of acquiring such information. These aspects are key due to their importance with the commercial world. If the benefits of analytics outweigh the cost to perform them then it is not a commercially viable route to follow.

A. Cost

1) Web Analytics

Compared to other forms of Big Data analytics, the price of web analytics is negligible. Most web analytics companies offer a free option which only caps visitor data per month. (Google Analytic Features, n.d.) This means the wide range of tools they offer is still available. Other solutions exist that offer various extra features to assist in web analytics, but most do not balance out with the fact they are not free. (Kaushik, 2010) When searching for a web analytics solution and the one you find is a paid service, it is suggested you ask the company the top 5 reasons their service should be chosen over a free equivalent to weigh up the benefits vs. the price for yourself.

Web analytics is usually done through a web interface, meaning no specialist software or hardware is necessary. Because of this, only a single machine is needed to interpret the data which cuts down the costs. As it is an entire analytics solution, the procurement of the data and its analysis are all done by the same company. This means the only costs involved
would be training staff to understand the results of the analytics to better improve the company as a whole.

2) Big Data

The costs of Big Data are split up between getting the data, processing the data and understanding the results. There are 2 main ways of getting Big Data: free public sources and data mining. Free public sources are commonplace throughout the internet, providing information on anything from geological surveys to the amount of internet capable users in Calgary. Due to coming from mostly research-based backgrounds, these types of datasets are useful for product or service research. Data mining lets you glimpse into the world of the consumer, through public Twitter and Facebook usage as well as forums, blogs and message boards. Attaining this data can either be done in-house or a pre-compiled data set can be purchased.

The data can either be analysed by the company or by a third-party. If the analysis is performed by the company, several key components are required. A network of sufficient processing power to make processing the data-set as efficient as possible is needed as well as a trained IT team to maintain the hardware and utilise software to process the data, such as Apache Hadoop. (Winter and Gilbert et al., 2013, p. 1-3) If the third party route is chosen, as stated above some companies provide solutions with the data included. In this case, the price would be as a package, instead of negotiated or pay-as-you-go like other cloud based processing services.

B. Data Produced and its Use

1) Web Analytics

The data produced by web analytics, as previously stated, consists of anonymous visitor information. This information comprises of where the visitor is, when they connected to the website, how they got there and where on the website they connected to. The data is then pre-organised into helpful graphs and figures that a marketing team can interpret easily without the need for a lot of extra training. Because of the focused nature of web analytics, the data is easier to process but less broad than the multitude of applications Big Data can offer. Certain web analytic services can be connected to other sources to provide deeper insights, for example Google Analytics which allow you to include search engine optimisation, finding the keywords that attract consumers. Other such connections are e-Commerce systems, to discover trends in popular online items.

2) Big Data

As the datasets used can be broad, so can the resulting data. This could consist of anything from a few percentages to an entire table of information. Because of this, the end result would need to be interpreted in a way which would be beneficial to the business instead of it being directly applicable. A positive side to this approach can be seen in the broad amount of answers available. Being able to see trends on a global scale can be very beneficial to a business assuming they can profit from it correctly.

C. Difficulty

1) Web Analytics

Because most companies offer it as an all-inclusive package, the end result tends to be as user-friendly as possible. The data collection and analysis is all performed automatically behind the scenes, making the only outcome viewed by the end user the results of the analysis. This is because it can be used by marketers as well as web developers with little training needed. (Peterson, 2004) As previously stated, the results received can be directly applied to the SME without having to be converted into a realistic goal due to the fact that the data is correlated from the SME.

2) Big Data

Big Data is not as user friendly due to the complexities involved both behind the scenes and in interpreting the data itself. Interpreting the data requires expertise that could take years to learn. In fact, trained Big Data professionals are in such high demand there is a shortage in the US of between 140,000 and 190,000 people. (McKinsey, n.d.) Unless outsourced, Big Data has a lot of hurdles before meaningful insights can be acquired, from getting the equipment necessary to training staff in maintaining and using it. Big Data solutions utilise technologies such as MapReduce to split the large task into smaller, more manageable jobs to be conducted by each node in the network. Due to this, custom algorithms need to be created to accommodate each different data type and required job. Unless outsourced, this requires trained analysts which, as mentioned previously, are in short supply. (Cohen et al., 2009)

V. Conclusion

As an SME, there are many approaches that can be taken to get a better insight in the business world whether through data on your company, your customers or the world in general. Before such ideals are pursued, questions such as budget and approach should be answered thoroughly.

The main question is which approach is best to invest the money into. With only using Web analytics as a base, as an SME the requirements would fit into the free category of most online analytic solutions. If the company was to expand into a larger format, the same insight may require a monetary investment through reaching over the free limit of web visits.

Another question is whether the business has a web presence worthy of being analysed. An SME that provides a local service without the need of an interactive website may find little benefit in web analytics whereas a business that wholly comprises of an e-Commerce solution could find the insights given invaluable.

If the SME may soon expand out of such a label, investing in in-house big data analytics could be worthwhile as pay-as-you-go services will end up costing more as the amount of data needing to be analysed grows. Finally, the question of training; Is it more worthwhile to train/hire a team to perform such analytics in house and maintain the equipment throughout, or outsource the entire job and instead have a single trained team to go through the results of the analysis to discover a solid approach for the company.

Although certain big data approaches will give great insight into the business world and even the internal running of your own SME, it may not be cost effective for an SME to aim for
them. Other forms of analysis exist within reach that are capable of providing effective results for a fraction of the cost and these forms, such as web analytics, could be all an SME needs to have a firm hold on their consumer base with the opportunity of increasing it. Although web analytics fails to provide the broad insight that big data can provide such as internal details within the company, the actionable results that it provides can more than make up for it, allowing a company to view the results from analytics and come up with a strategy tailored to their needs.

VI. REFERENCES


Abstract—SME’s (Small and Medium Enterprises) face a big challenge in competing with large established enterprises in their respective market. SME’s are being advised to adopt Big Data technology as a solution towards cutting the gap between themselves and the larger enterprises, having market edge over rivals and the chance to be more established in the industry. This article focuses on a number of issues surrounding enterprises that have implemented Big Data technology while also advising SME’s to abstain from the technology until certain laws and regulations are in place to protect the privacy of consumers.

I. INTRODUCTION

Big data is gaining more attention and popularity each passing day. While most of the emphasis is on the benefits, opportunities and success stories in general, a small minority are more focused on bringing to light the ugly side of this much talked about topic including various ways we are being monitored and manipulated by enterprises and governments. This article focuses on the downsides of SME’s adopting this technology and reveals the various ways internet users have been leveraged due to misinformation and lack of transparency. This article will discuss the importance of maintaining our choices for what we wish to keep private, why SME’s should stay clear of this technology for the time being including ethical concerns surrounding the use of Big Data along with various case studies, and suggestions on how big data and privacy coexist. Consumers need to decide if their privacy is worth being traded for unguaranteed security and the claim of “free” customized services (which we are unknowingly being charged).

II. BIG DATA

The amount of data in our world is growing astronomically in ways that pose a challenge to organizations. With the availability of the internet, different devices serve as data collection sources; ranging from digital sensors in equipment to mobile phones, surveillance cameras, human and automobile trackers etc. The amount of data collected and transmitted through these mediums can only be overwhelming to store and expensive for organizations to manage. But with the help of technology and skilled professionals, organizations are able to analyze this massive dataset to discover trends and patterns that can help businesses make more informed data driven decisions. Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured (SAS, 2014). Two definitions of Big Data that encompasses its purpose and features are; Big Data is a collection of variety of data from traditional and digital sources in and outside an organization that is too large for conventional database management tools to capture, store and process, and also represents a source for ongoing discovery and analysis, and Big Data refers to the practice of companies collecting millions of facts about customers and using those facts to predict trends and develop better sales and marketing strategies (Payton & Claypoole, 2014). In essence, mobile devices make it easy to collect data, while big data capabilities make it increasingly trivial to take the resulting mass of supposedly anonymized data and tease out the kind of specificity that the anonymizers were trying to erase (Meyer, 2013). Large enterprises would argue that the main objective for adopting this new technology is to serve its customers better through personalized services, but in reality, these enterprises analyze the very sensitive and private information they gather about their customers with the sole aim of predicting buying habits and influencing spending decisions.

III. CONSUMER’S DATA

The deeper technology becomes embedded into our lives, the more it threatens our privacy (Payton & Claypoole, 2014). Most of the computing devices available today are interconnected, a platform for devices to gather users’ data, share these data and transmit the data collected into large databases managed by businesses and governments. Age, gender, education, employment, marital status, religion, finances, sexuality, location etc. include some of the personal details that are accessed and transmitted by these smart devices without our conscious consent. What modern data science is finding is that nearly any type of data can be used, much like a finger print, to identify the person that created it: your choice of movies on Netflix, the location signals emitted by your cell phone, even your pattern of walking as recorded by a surveillance camera (Tucker, 2013). In effect, the more data there is, the less any of it can be said to be private, since the richness of that data makes pinpointing people “algorithmically possible” says Princeton University computer scientist Arvin Narayanan (Tucker, 2013). Large technological enterprises claim to offer free online services but yet leave behind cookies (text files) in users’ devices capable of gathering and storing personal information, which is then returned to the respective
server the next time the page is referenced. It is fair to say, the limited disclosure of personal data usage (or no disclosure at all) by organizations usually creates distrust among consumers (Bertolucci, 2013). “Consumers get the most irritated when [their data] has been collected in stealth, and they are not very clear on how it’s being used and, more importantly, how they’re benefitting from it”. The result of this damming effort is the victimization of consumers through targeted and behavioral marketing.

IV. PRIVACY

Privacy is the control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others (University of California, Irving Office of Research, 2014). Privacy is a right every individual should possess, the first line of defense to our possessions, the perception of being in control of the access that others have to us. Privacy should exist as a gateway between internet users and the cyber world to establish trust, transparency and an essence of freedom. The perception of privacy should not always be confined to secrecy. Privacy is not always about hiding something, but more about personal control and freedom of choice. Privacy is crucial to protect and support the many freedoms of responsibility that we possess in a democracy but sadly, society has reached a point where technology is advancing at a speed that’s too quick for the law to keep up to. A good example is the story of Target (the second largest merchandise retailer in America), using their pregnancy prediction methods supported by the application of Big Data, to foretell a teenage girl’s pregnancy even before her father had any knowledge of it, by using surveillance provided by the big data platform to totally disregard the right to privacy of both the individual and the family. According to code 1.7 of the Association of Computing Machinery Code of Ethics, Respect the privacy of others. This code states it is the responsibility of professionals to maintain the privacy and integrity of data describing individuals. This includes taking precautions to ensure the accuracy of data, as well as protecting it from unauthorized access or accidental disclosure to inappropriate individuals (ACM, 1992). Even worse is the rate at which the government and society changes their definition of privacy intrusion which can get really tricky and difficult to staying on the right side of the law and society standards. Three areas society needs to put under the microscope concerning big data and privacy are surveillance, disclosure, discrimination.

A. Surveillance

Surveillance is the antithesis of privacy (Cavoukian, 2013). In this modern society, surveillance is everywhere and creates a paranoid feeling within individuals. The feeling of surveillance can arise from the mere collection of information, as when visits to sensitive websites are tracked. It menaces our intellectual privacy and it gives the watcher a power advantage over the watched, which can be used for blackmail, persuasion, or discrimination (Washington University, 2013). The reduction of surveillance in our society will aid the rebuilding of trust and the construction of a more privacy friendly internet. Surveillance minimization requires surveillance to be targeted rather than universal, controlled and warranted at the point of data gathering rather than of data access, and performed for the minimum necessary time on the minimum necessary people (University of East Anglia, 2014). It has a profound effect not only on individual privacy but also on freedom of association, freedom of expression, assembly to protection of discrimination, and human rights. An ideal instance of how surveillance is used to violate basic human right as far as privacy is concerned is the case of the NSA’s previously undisclosed program PRISM, which allowed officials to collect material including search history, the content of emails, file transfers and live chats from Facebook, Microsoft, Google, Yahoo, Apple and other US internet giants, while also monitoring or recording presumably every phone call in the United States. With this program, the NSA was able to reach directly into the servers of the participating companies and obtain both stored communications as well as real-time communication on targeted users (Glenn, 2013). After seven weeks of steady media coverage, the percentage of Internet users worried about their online privacy jumped 19 percent, from 48 percent in June (when the story first appeared in The Guardian and Washington Post) to 57 percent in July, according to Annalect, Omnicom Media Group’s data and analytics company (Bachman, 2013). If normal citizens aren’t declared or perceived as a threat to society and are being monitored, then unknowingly, they are participants of a research with certain objectives. Not seeking the consent of those individuals raises serious ethical concerns and is a serious violation of privacy, and not compensating those participants is just blatant oppression.

B. Disclosure

Disclosure of data outside of the context in which it was collected has also emerged as a critical concern in the ever growing world of big data. The trend towards increasing data disclosure has begun with the advent of the “open data movement,” which has demonstrated the benefits that could be derived from the public disclosure of public sector information (Davies, 2010). Enhancing the legibility and comprehensibility of such data has been aided by the development of various tools for data analysis and visualization, primarily encouraging accountability and transparency in the public sector. The private sector has adopted these tools, using them mostly for commercial purposes with the aim of using analysis of customer preferences and behavioral data to boost their marketing and sales strategy. In this regard, big data analysis has shown that significant value can be extracted from the collection and aggregation of data into large datasets so as to be able to better identify patterns and correlations amongst them (Lohr, 2012). Public disclosure of personal information is increasingly becoming a trend in the world today as more people continue to profit and take advantage of either sharing or exchanging personal data. The recent growth in popularity of activity trackers (such as the Fitbit, Basis, or Nike’s Fuelband) clearly illustrates that users are becoming increasingly comfortable with disclosing personal information (including health data) in exchange of highly personalized services tailored to their own preferences and needs (Golle,
2006). Yet, with these devices, users are not just collecting data about themselves; they rely on third party operators to process a large amount of personal information coming from many different sources (Richards & King, 2013), and to analyze it by means of statistics and sophisticated data analysis techniques in order to extract new information that could not be easily inferred from the individual dataset of any single person (Cambria & al., 2013). While this is likely to impinge upon the individual right to privacy and data protection (Craig & Ludloff, 2011), in the case of many online services, consent is obtained by means of long and intricate Terms of Service (ToS) which users necessarily have to agree to in order to benefit from the services offered by online operators (Bradshaw, et al., 2011). Most users do in fact, generally agree to their personal data being collected and processed by third party online operators in order to benefit from a more customized service that would not be possible otherwise (Oboler, et al., 2012). Social media giants Facebook have been accused of selling user access to companies, enabling these companies to track users on Facebook using certain customer details, giving retailers the opportunity to compare their databases with information on Facebook. The social networking site is allowing companies to trawl through its 900 million users looking for email addresses and phone numbers so it can better target adverts (Bates, 2012). Consumers who have handed over some form of personal data when shopping online will be seriously affected as that company can track you down on Facebook for more targeted marketing.

C. Discrimination

Greater knowledge of customers creates new potential and power to discriminate (Trader, 2014). Some people think that big data is really quite fantastic because you’re working at a mass level and therefore you can’t actually conduct group-based discrimination. It’s actually quite the opposite. Big data is not color blind, it’s not gender blind and, in fact, marketers are using big data to have ever-more precise categories about you (Pressman, 2013). According to section 7.8 of the IEEE Code of Ethics “To treat fairly all persons regardless of such factors as race, religion, gender, disability, age, or national origin” (IEEE, 2013), is supposed to serve as a means of protecting people. Regardless of organizations admitting to discrimination or not, big data gathers much information that it presents adopters an opportunity to discriminate. It is a possibility that enterprises would use data collected using big data to charge customers differently in the future. First degree price discrimination involves charging every individual customer a price based on their individual willingness to pay (Ozimek, 2013). Ozimek goes on to state “as big data and online buying increases the information that businesses have on us, the ease and profitability of first degree price discrimination will become difficult to resist” (Ozimek, 2013). “Also, a lender for example, who didn’t want borrowers of a certain race could show online offers only to people whose social network activity fit certain parameters. Banks must report detailed statistics about their actual lending activities to regulators, but web advertising parameters are seemingly free of discrimination. By never putting offers in front of unwanted groups, and thus never formally rejecting them, those who engage in online discrimination could sidestep fair lending and redlining laws that apply in the physical world 20°. The mixture of Big Data and social media also poses a problem as users tend to publicly post various things online, being naïve to the ability of data mining services possessing the ability to use this information without the individual’s knowledge. A recent study at Cambridge University looking at almost 60,000 people’s Facebook “likes” was able to predict with high degrees of accuracy their gender, race, sexual orientation and even a tendency to drink excessively (Pressman, 2013).

V. SUGGESTIONS ON HOW BIG DATA AND PRIVACY CAN COEXIST

Although very difficult to find a common ground for the coexistence of Big Data and privacy in today’s technological dependent society, it is at least feasible that these two much talked about topics can find a way to rely on one another for survival. Listed, are few suggestions that can lead to a new era for Big Data:

A. Monetary Benefits

Evidently, big data implementing enterprises need our data for them to function at optimum efficiency. Customized tailored services are no longer enough for users’ whose data is acted upon to make massive profits. If companies can’t respect users’ privacy and rely heavily on user data, they should be willing to part with a certain amount of revenue to consumers whose data are being used.

B. Paywall

Platform providers could introduce contracts that rather than paying for access, allows users to pay for privacy, simply by allowing users to choose the kind of data they are willing to relinquish to advertisers and data they are not willing to disclose. This will serve as an honest approach in dealing with privacy concerns of consumers, and it highlights an enterprise’s stance on transparency.

C. Boycott Services

Activists need to unite, spread awareness on how our personal data, aided by society’s naivety and negligence is being leveraged, and challenge these internet giants such as Facebook, Twitter etc. by boycotting their services as they did to Stop Online Piracy Act (SOPA).

D. Right to be forgotten

This will require the data controller to take all reasonable steps to have individuals’ data erased, including by third parties ‘without delay’, for the personal data that the controller has made public ‘without legal justification’ (Data Guidance, 2013).

E. Transparency and Honesty

SME’s willing to adopt big data can be honest with the public from the get go. SME’s who are open about their data
gathering techniques and data use will experience the least consumer revolt.

VI. CONCLUSION

The implementation of Big Data seems all the more tempting in recent as data gathering is aided by the sheer availability of all sorts of technological devices with the ability to amass and transmit data. Also with the growing reliability on cloud computing tools and services, storage and analysis of data collected seems less overwhelming than in preceding years. But SME’s need to look past the various benefits that can be gained from joining the Big Data trend set by large enterprises. Just as evidence is considered inadmissible in the court of law if it was achieved illegally or wrongly, SME’s should view certain benefits in the same manner as the way which data was gathered and used in most cases of Big Data usage and implementation is considered morally and ethically wrong, and even illegal in some cases. Make no mistake; this isn’t a condemnation of the existence of big data but rather, an opportunity to address its implementation and effects towards society. Until Big Data is implemented in the right way and privacy principles are in place that give enterprises no choice but to respect the right to privacy of individuals, then SME’s should put on a brave act and turn the blind side to the use of Big Data.

VII. REFERENCES


Big Data Breaches
What Are The Security Risks And Consequences Associated With Big Data?

Samuel James
School of Computing and Mathematics
University of Derby
Derby, Derbyshire
S.James5@unimail.derby.ac.uk

Abstract—This report looks at the security concerns and considerations surrounding big data analytics. The report is broken down into three sections outlining the current concerns and considerations for Big Data. To determine the likelihood of an attack this report looks at the increasing trends found within various data studies for organizations that are experiencing an increasing number of attacks as well as increased costs for dealing with the breaches. The report concludes with the importance of investing in adequate security in order to minimize the cost and reputational damage of a data breach.

Keywords—Big data breach, consequences of a data breach, data breach risks, reputational consequences, big data concerns considerations.

I. INTRODUCTION

Data privacy has always been a key concern within the technological world. As technology evolved so did our power to store more and more information into smaller and smaller spaces. This concern for security has always shown to be justified as there has been a long history of news stories involving computer hacking or theft of sensitive and confidential information over the last decade.

One of the more recent trends from the computing world has been that of big data analytics. Big data analytics is simply processing extremely large amounts of data to uncover patterns and correlations that can be used help organizations make better business decisions, which in most cases means more money (Rouse, 2012). There has always been a lot of concern surrounding the collection and storage of such large amounts of data such as the ethical implications but in this report this report looks at the issues raised from a security point of view as well as looking at the security risks that threaten big data organizations and the consequences that can impact a company in the event of a data breach (Robinson, 2014).

II. CONCERNS FOR IMPLEMENTING BIG DATA

Big Data has been around since the first decade of the 21st century and has been embraced by many big organizations, especially online and startup firms such as Facebook, Google and eBay which were able to build their analytics and IT infrastructure around Big Data technologies in such a way that they could utilize all the benefits big data has to offer (Davenport et al, 2013: p.1). Well-established organizations that are looking to implement big data analytics however must find a way to integrate it with everything else that the organization does such as finding a suitable way for the big data analytics to coexist with the analytics of other types of data (Davenport et al, 2013: p.1).

There are a number of other concerns to take into account however when considering the implementation and use of big data analytics, one of the biggest concerns being security. Because Big Data environments allow organizations to aggregate more and more data, most of which is financial, personal, intellectual or some other form of sensitive data they become a prime target for hackers. One of the most recent and notable examples of a big data security breach was the December 2013 Target Corporation retail chain who reported that cyber-criminals had access to information associated with 70 million customer accounts including credit and debit-card information as well as customer names and addresses (Kolakowski, 2013).

To try and help raise the security awareness for big data environments IBM published an article in October 2012 detailing the unique challenges of securing big data environments along with security fundamentals and solutions to the issues highlighted. Some of the unique challenges outlined within the article for organizations to come to terms with are:

"Schema-less distributed environments, where data from multiple sources can be joined and aggregated in arbitrary ways, make it challenging to establish access controls
The nature of big data- high volume, variety and velocity-makes it difficult to ensure data integrity
Aggregation of data from across the enterprise means sensitive data is in a repository
Big data repositories present another data source to secure and most existing data security and compliance approaches will not scale" (IBM, 2012: p.5).

Big Data environments have currently only been deployed in major ways at some of the most innovative organizations in the world which creates a window of opportunity to establish a set of best security practices (IBM, 2012: p.6).

III. CONSIDERATIONS FOR BIG DATA SECURITY

As data volumes increase so does the risk of greater security breaches. In order to meet government compliancy organizations need to ensure they have mechanisms in place to secure the data (Advantech, 2013: p.2). As IBM (2012: p.6) explains Protecting data should take a holistic approach in
order to protect organizations from a complex threats across variety of systems due to the fact that protecting data security is a key responsibility which should be part of every best practice (IBM, 2012: p.6).

By planning big data security immediately, organizations can address data security in an effective and efficient manner. One fundamental method for achieving this encryption technology on both a software and hardware level that can operate on selected data during executions or across an entire disk (Advantech, 2013: p.2). Whilst encryption is a fundamental method for securing data, organizations must take into account that software-based encryption will add significant strain or load on a database server's CPU which increases the cost and complexity, especially in situations where the solution is required to scale (Advantech, 2013: p.2). By using best practices and building security into big data environments organizations will be able reduce not only the risks but also the costs and deployment pains (IBM, 2012: p.7).

There are a variety of products that offer hardware-based encryption such as Intels Communications Chipset 89xx Series which can help significantly boost performance as well as achieving significant cost savings (Advantech, 2013: p.2).

One way of achieving fast and reliable analysis of big data is by using Hadoop-based systems. This is because Hadoop-based systems can be deployed along side existing database systems, which allows them to combine traditional structured data and new unstructured data sets in powerful ways (IBM, 2012: p.7). Hadoop works by using a high-performance parallel data processing technique known as MapReduce along with its own column-orientated database management system known as the HBase which runs on top of the Hadoop Distributed File System (HDFS).

IBM (2012: p.8) once again listed some security strategies that can be implemented to secure Hadoop environments such as:

- Sensitive data discovery and classification
- Data access and change controls
- Real-time data activity monitoring and auditing
- Data protection
- Data loss prevention
- Vulnerability management
- Compliance management

Organizations should also be prepared to follow audit requirements as compliance mandates are enforced in big data environments as they are in traditional data management architectures (See Figure 1).

In an effort to help security and risk professionals a report was released by Forrester Research, Inc that created a framework by breaking the problems down into three areas (See Figure 2).

By following this framework organizations can control and secure the extreme volumes of data in big data environments (Forrester Research, Inc, 2012). The breakdown of the framework as IBM (2012: p.10) states that “the unfortunate reality is that organizations need to adopt a zero trust policy to ensure complete protection.”

In order to better secure big data environments Forrester Research, Inc (2012) actions that organizations can take include:

- Moving controls closer to the data
- Utilizing existing technologies to protect and control big data
- Clearly defining policies for data archiving and disposal
- Controlling access to big data resources
- Monitoring user behaviour

One method to securing these environments such as Hadoop is using a technology such as IBMs InfoSphere Guardium solutions, as they scale to protect both big data environments and traditional data management architectures as well as protecting against threats such as insider fraud, unauthorized changes as well as external attacks (IBM, 2012: p.12). Unfortunately however, due their nature and value, big data environments will always be a prime target for cyber-criminals. Organizations should therefore be prepared for the consequences of a big data breach if they decide to collect, store or analyze big data.
IV. LIKELIHOOD OF A SECURITY BREACH

As IBM (2012: p.4) outlines the pressure to make quick decisions can result in data security professionals being left out of important decisions or be viewed as inhibiting business growth. This unfortunately can lead to lax in data security and therefore a data breach which can have a number of negative impacts upon an organization.

There have been a number of technical reports that look at the root causes and trends, as well as the consequences, of a big data breach, conducted by organizations such as the Ponemon Institute and PricewaterhouseCoopers (PWC).

The Information Security Breaches Survey 2013, conducted by PWC, showed that the number of security breaches affecting UK business continued to increase from 2012 as well as the fact that small business are now experiencing incident levels previously only seen in large organizations (PWC, 2013: p.2). Security breaches on the whole for any sized organization were also shown to of increased since 2012 as well as showing sizeable increases in the average number of breaches in the year, the cost of the worst breach of the year and the overall cost of security breaches.

Overall the survey showed that 93% of large organizations (over 250 staff) had a security breach of some sort and 87% of small business (less than 50 staff) had a security breach of some sort which was up from 76% the previous year (PWC, 2012: p.2).

In regards to the frequency of security breaches, the PWC technical report showed roughly 50% more breaches than the previous year. The median number of attacks for small businesses for the 2013 survey was 17, up from 11 the previous year, and 113 for large organizations, up from 71 the previous year (PWC, 2013: p.2).

V. TYPES OF BREACHES

Unfortunately organizations are under threat from both external and internal sources, though the majority of breaches are due to outside attacks from criminals, hactivists or competitors. Whilst smaller businesses do not face as many attacks as large organizations over the course of a year, a large business will face a significant attack every few days on average (PWC, 2012: p.2).

The survey showed that 63% of small businesses were attacked by unauthorized outsiders in the year which is up from 41% the previous year (PWC, 2012: p.2). 15% of small businesses also detected that an external source had successfully penetrated their network, up from 7% the previous year, and that 9% of small businesses knew that their intellectual property or confidential data had been stolen, up from 4% the previous year. Denial-of-service attacks on small businesses were also shown to of increased from 4% to 9% over the year (PWC, 2012: p.2). This is a substantial increase in the number of attacks aimed at smaller businesses, highlighting the potential value of sensitive data to outsiders. It also demonstrates the importance of securing data within these environments.

The number of security breaches due to internal sources had also increased within small businesses with 36% of the worst breaches in the year being caused by inadvertent human error as well as another 10% due to deliberate misuse of systems by staff. 57% of small businesses overall suffered a staff-related security breach, up from 45% the previous year, and 17% knew that staff had broken their data protection regulations, up from 11% the previous year (PWC, 2012: p.2).

VI. CONSEQUENCES OF A BIG DATA BREACH

Any organization looking to reap the rewards of big data analytics should always be prepared for a data breach that involves the loss or theft of thousands of records. These big data breaches can have a number of consequences for any organization. In the 2011 Cost of Data Breach Study conducted by the Ponemon Institute (2012: p.2) the average organizational cost of a data breach is $5.5 million which averages out to $194 per compromised record.

The unfortunate reality however, is that there are many other consequences that an organization may face after a breach. Another report, conducted by the Ponemon Institute, surveyed the possible responses following a material data breach. The survey looked at the responses organizations expected and were concerned with the most following a material data breach as well as cases where the response had already happened. The results showed that organizations mostly feared loss of customers and business partners as well as negative public opinion and media reports (See Figure 3).

![Figure 3. Table Showing the Likely Responses Following A Material Data Breach (Ponemon Institute, 2013).](https://example.com/figure3.png)

It is also worth noting that the "Serious financial consequences" response can come in a variety of forms. Organizations may have to pay for costs that the bank incurs dealing with a data breach, losses suffered by customers as well as costs relating to improving data security. A good example of such a scenario can be linked back to the 2013 Target Corporation big data breach where, according to Reuters, (Finkle et al, 2014) Target shares dropped 11% and net profit for the crucial holiday quarter were down 46%.

As the survey shows aside from the financial consequences there are a number of other consequences organizations should be prepared to face in the event of a data breach.
Big Data breaches also pose reputational and image damage to organizations. According to the Ponemon Institute (2011) the value of a an organizations brand or image can drop by as much as 17% to over 31% in the event of a data breach. 54% of organizations also believe that it can take 10 months to over 2 years to restore a company reputation following a customer data breach (Ponemon Institute, 2011).

VII. CONCLUSION

Whilst big data can be very valuable to an organization the consequences of a big data breach can have a brutal impact on an organizations image as well as their future. The findings of the technical reports, shown within this document, prove that confidential data will always be a target for theft. Any organization, large or small, looking to reap the rewards from big data analytics needs to be prepared to invest in adequate security as well as prepare for the worst case scenario. By doing this organizations can try to minimize the negative impact of a breach whilst at the same time utilizing big data analytics to its full potential.

VIII. REFERENCES


Can SME’s Use Big Data To Help Make A Successful Game?

Michael King  
Faculty of Business, Computing and Law  
University of Derby  
Derby  
100239972  
M.King4@unimail.derby.ac.uk

Abstract—it’s difficult for an SME games company in a world where game success rates are as low as 20%. Big data has the opportunity to increase this success, through enhancing the game development process, improving a gamer’s experience and providing target marketing insights. Unfortunately, the challenges of value, validity, veracity, volume, volatility and vitality are standing between these opportunities. This report will explore these opportunities and challenges to see if big data can really increase a games success rate.

Index Terms— big data, SME, game, value, vitality (key words)

I. INTRODUCTION

In a world where digital content is everywhere, it is exciting being a game’s SME in a time where games are the fastest form of mass media (Cross, 2011). Making a game is usually a long process, requiring plenty of creative designers, skilled developers and the money to fund a project (Brownsword, 2009). Even with all of these in place the failure rate of SME game projects is over 50%, with only 20% of the games to hit the market making a profit (Plunkett, 2008, UOK, 2011, Williams, 2012). With this in mind, is there a solution to building a more successful game? The answer could lie in using big data.

So what is big data? Big data is a collection of data from multiple sources (digital and traditional) which has the potential for future analysis, in order to find useful trends (Arthur, 2013). It is used in multiple industry sectors worldwide (Tien, 2013) for a variety of opportunities and benefits. Although in order to get these, plenty of challenges have to be overcome. This report will discuss these opportunities, benefits and challenges, which when used correctly, could lead to a more successful game.

II. OPPORTUNITIES AND BENEFITS

With 2 billion players worldwide generating a huge volume of over 50tb of data per day (Rijmenam, 2013), there are plenty of opportunities to collect and use this data. These opportunities come from collecting or using existing online and offline data (Bertolucci, 2013), which lead to benefits such as more successful game development, improved online player experience and finally useful target marketing (Chulis, 2012). The following sections will explain this in detail, providing an insight to what big data can bring to an SME.

A. Game development process

1) Using existing big data

The first opportunity for SME’s is to plan a game based on existing big data. This is smart as it allows companies to see all sorts of useful information including current market trends, gamer habits and popular genres on different platforms (VandenBerghe, 2013). This can lead to better game development as games are empathically designed based on what customers currently want and what works well (Sturt, 2014). For example, an SME has an idea for a new puzzle game but doesn’t know which platform is best to develop for. Using existing big data, the company could see that currently puzzle games are more popular on tablets/smartphones (ESA, 2014) and are played by social gamers more than traditional gamers (Farago, 2011, ESA, 2014). As a result, the game would be best developed for smart devices, with a target audience of social gamers. Be aware though as big data is Volatile and could change at any time (Normandeau, 2013, LA, 2013), with one trend making all existing data useless (see section III). Another trap to avoid would be overreliance on big data, turning games more data driven than creative driven (Fleck, 2014). Gamers want a unique, exciting and fun experience, not the same game over and over (Thompson, 2012). Games which are purely data driven will be boring and uncreative (Voller, 2012), reducing success. Big data should be used as an aid, not a replacement.

2) Collecting own big data

The second opportunity to discuss is the use of achievements/trophies to enhancing level design and gameplay. Achievements (Xbox consoles) and Trophies (PlayStation consoles) are used to reward players when they complete a certain task or level in a game (Norman, 2009, GB, 2013, Gotto, 2013). The player can keep a track of these from multiple games (which include them), and compare with friends, enhancing the gaming experience (Microsoft, 2013, Sony, n.d.). The popularity of these by players has increased so much that Apple and Android are encouraging developers to include them in their games, creating a set of guidelines to do so (Apple, 2013, Google, 2014).
Since these trophies are in most games, why not use them to collect big data and to analyze game features and gaming habits? For example, take the difficulty factor. Design a set of trophies to award a player for completing each difficulty setting of a game (easy, medium or hard). When a player has achieved a trophy, the game can be synced to the companies’ cloud servers (since cloud solves storage issues (Ningyuxin, 2013)), instantly or when the player signs in online (Wei-Tek, 2013, Google, n.d.). As a result, huge data volumes can be collected from millions of gamers and analyzed to see which game difficulty is the most popular. This can be applied to the design of any trophy/achievement, in order to collect data on any task/level in the game.

Another useful example could be to see how long the average player took to complete the game (VandenBerghe, 2013). Trophy data includes data/time information, which algorithms could be created in order to compare when the first level completion trophy was earned and when the last one was earned, providing player completion time. This is also useful as it could show if a player finished the game or not. If the data shows that a high percentage of players stopped playing after a certain level, then that level design will need to be rethought. This can be done in the current game online data, by comparing when the first online patch/update (Makuch, 2014b) or in the next game. Since only 10% of players ever finish a game (Snow, 2011), this big data method could provide a useful insight to why this is happening to companies worldwide.

Correctly designing these trophies can lead to a better development process of a company’s game, especially since it’s based on big data which is valuable (tailored to company’s game) and has high validity/veracity (comes directly from the gamer and the game, making it a trustworthy source).

B. Improved Gamer Experience Online

With the average household owning at least one console to play games on, and 62% of them playing games online, it is becoming increasingly important to have a stable online service for millions of gamers to enjoy (ESA, 2014). A lot of companies have inaccurately predicted online gaming demand, with huge volumes of players causing their servers to crash. The delay in solving this problem leaves gamers without a service, decreasing their satisfaction along with the company’s reputation (Gauder, 2013). It also risks gamers leaving to play other online games whilst fixing the issue, with some never returning. A recent example of this is the Pokébank cloud service, which allows players to store and access 3000 Pokemon online (Makuch, 2014b). The service crashed within hours of going live due to unpredicted demand, and was delayed for 2 months, frustrating fans globally (Akash, 2014, Alan, 2014, Makuch, 2014b). Big data can help predict these demands/peaks in activity, in order to prevent these crashes. This can be done by collecting big data on daily, weekly and monthly active users (Johnson, 2011). Knowing these gamer habits will provide useful times of when players are most likely to be playing the game online. Existing big data collected by other companies could be also analysed. If most company’s data show that their server peaks are between certain times of the day, then most likely this will be the case for all (Ahamed, 2013). Be wary though as big data from USA game companies are in different time zones than UK ones, making the data validity different depending on country.

C. Target Market Opportunities

This final opportunity is to recommend how SME’s could collect/use big data in order to target products to gamers. Big data is widely used to target market in multiple industry sectors (Dunay, 2013, Hemsley, 2013, Radermecher, 2014). Big data is collected in multiple ways including reward schemes and free products (Arnett, 2013, TMD, 2013). An example of this would be Tesco’s Clubcard reward scheme. Customers gain points towards money off vouchers for scanning their Clubcard during every purchase. The customer saves money and stays loyal to the company, whilst Tesco gets valuable data on purchasing habits. Tesco then sends out target adverts to their customers, based on what they purchased (Ruddick, 2014).

Any games company could use a reward scheme approach to collect big data from gamers. For example, SME’s could offer gamers free digital content by filling in a short online survey (Wasserman, 2013). Most players will be so focused with the exciting free content that they won’t mind filling it in. Examples of free content could be downloadable characters, costumes, accessories or even levels. A similar approach is used by Facebook and mobile developers, who collect huge volumes of big data by allowing players to play free online games (Smith, 2012, BDS, 2014).

These approaches are useful but which data should be collected in order to target gamers successfully? This is important as value is a critical factor when it comes to big data (Biehn, 2013), as there’s no point collecting data unless it is going to be useful to the company. An example of useful data would be to see if a gamer is a daily, weekly or monthly user of the game (Johnson, 2011). This provides a useful insight as gamers who are active everyday are more likely to be interested in purchasing additional content compared to monthly active users. These daily users can be targeted new content as a result. Other useful data could be demographics (Rogers, 2012), in order to see which age group or location is playing the game the most. If a certain location is playing the game more, set up a special event there to talk to fans to see what they like and to also promote new gaming content. Targeting players who regularly play will increase the vitality of the game, extending its life and making it more successful.

There are plenty of opportunities to collect useful data but remember to respect gamers. Sending promotions regularly will frustrate them (Fletcher, 2013), negatively affecting the company.

III. CHALLENGES AND RISKS

The above section shows that big data can lead to designing a better game, improving the online experience and target marketing players. These are all useful factors which can lead to greater success of a game. In order to get there several challenges and risks need to be overcome, which lie in multiple stages of the big data process (collecting, analysing, interpreting, decision making). This section will explore these
in more detail, to warn SME’s of common mistakes which can easily be made.

A. Challenge #1: Precious Value

The first challenge to overcome is deciding which data will bring the most value to the game. This is one of the most important factors as if data isn’t going to bring value, there is a risk of wasting resources (time, money, staff) collecting and analysing it (Tallon, 2013, Hardiman, 2013). For example, why collect data on every game platform when your company specializes in smart phone games? Taking this further, why collect data from traditional gamers when your game is targeted towards social players? Be smart before rushing out to collect data, ensuring that all of the approaches used are carefully designed to bring useful results (Miller, 2013).

B. Challenge #2: Uncertain Validity

The next challenge is that of validity, meaning is the data correct and justified? The key to understanding this is that correlation does not mean cause (Bort, 2013, Press, 2013). A good example of this would be the UK PC games market. For years analysts have predicted that it will soon die, yet in 2013 it made half a billion pounds (Dring, 2014). In this case, choosing not to develop for PC based on analyst’s data could have cost the company a valuable market opportunity. According to an IBM analyst, by 2015 80% of all big data will be uncertain (Eastonnn 2013). As a result SME’s would be smart to question any data trends before making decisions.

C. Challenge #3: Suspicious Veracity

The third challenge is veracity, which is all about if a company can trust big data. There are many factors involved before trusting it. For example is the data legal? If the company hasn’t had consent off of players before collecting the data then it would be breaching the data protection act (1998) and is liable for prosecution. Respect gamers and ensure they agree with the terms before collecting data and ensure that the data is anonymised also (Treacy, 2012). The second part of trust would be consistency (Benati, 2013). Is there wide enough evidence to suggest that the trends discovered are true? Is this trend true across the target market which the game is designed for? If the answer is no, then it should not be trusted (Mckendrick, 2014). The final trust issue would be the source of the big data (Shacklett, 2014). Is the data from a quality source? This is especially important if the company is using third party data. Always use a provider that is well known and respected, otherwise there’s a risk that important game design choices could be based on lies, reducing the success rate of the game.

D. Challenge #4: Incredible Variety

The next challenge is managing variety. With millions of data available out there to collect, sometimes a company can be overwhelmed with choice. Even if the data collected is all valuable at some level, it is important to organise resources in a way that focuses on the main goals first (Hardiman, 2013). If there isn’t a solid plan, it is easy to get distracted with so many exciting trends (Deloitte, 2011), and waste valuable employee time and company money investigating these trends further. It is good practice to focus resources on one trend effectively, than spend multiple on several trends chaotically. Variety isn’t all about the big data itself but also the providers of it. The Gartner Emerging Technologies Hype Cycle (Lehong, 2012) predicts that big data will be at its peak in the next 1 – 5 years, meaning that providers will be popping up left, right and centre. If providers are used it is important to find one that is right for the company, not just based on price, otherwise it risks providing data which isn’t relevant or valuable to the game project (Tallon, 2013).

E. Challenge #5: Huge Volume

This challenge is all about avoiding the risk of volume. As mentioned above, when choosing a big data provider, it is important to find one which is right for your company. Some providers will have solid advertising, stating that they have petabytes worth of data to offer. Although this may be true, (since 90% of worldwide data came from the past 2 years (Humetov, 2012)), only 5% of data is usually relevant to a company (Grimes, 2013). Ironically, it’s about finding the small data in big data (Oracle, 2013), with most big data in fact being useless to a project. It is better to have 1GB of data on gamer habits, than 1 petabyte of data on non-gamers. Finding this 5% will lead to a more successful game as trends will be relevant, helping with the opportunities above.

F. Challenge #6: Rapid Volatility

This challenge is one of the hardest to predict and adjust to. Volatility is all about the change of data, which is usually unpredictable (Ramalingam, 2013). A company could collect years’ worth of solid valuable data on gamer’s favourite platforms, genres and interactions. The next day a product could be released which can make all of the data collected invalid. A recent example of this would be smartphone technology. The release of Apple’s iPhone completely changed the games industry, shifting the market from traditional console gamers to social mobile gamers (Farago, 2011, Minkley, 2012). Popular handheld consoles such as Nintendo’s 3DS and Sony’s PS Vita have lost sales and customers to these devices (Fahmy, 2014). The price people pay for games also shifted, with many iPhone games being free, with money made on advertising alone (Rubens, 2014). Successful games can now be coded in a week by an individual developer, and make more money than a team worth of developers spending years on a project. An example of this would be Flappy Birds (Rubens, 2014). All of this change because of one trend, the iPhone.

With new technology products being released every 6 months, this rapid change could become even more frequent. This is especially worrying for games companies that could spend years on a project, only to find out on completion that their game isn’t popular anymore. Most small company’s futures are riding on their game release (Starkley, 2012), so this volatility could cause them bankruptcy. This leaves the question “is big data worth investing in due to volatility?”. Since unexpected trends are difficult to predict, it’s up to the games company if the risk of investing is worth it or not, but they should do so only if they can afford to (Tallon, 2013).
G. Challenge #7: Sustainable Vitality

The final challenge is vitality, keeping the game alive and profitable as long as possible. If a game isn’t sustainable, it is less likely to be more successful (Cookson, 2013). New content needs to be added in order to keep players engagement high (GB, 2013b). As a result, big data needs to be used as an iterative design process (Nielsen, 1993). This means that it doesn’t end when the product is released. More collection and analysis needs to be done, in order to see how the game can be improved and what players are still playing so they can be targeted additional content. This applies to every type of game on every platform. If a game is successful then hundreds of companies/developers will start looking and adding recommendations from reviews. As a result, consistency with big data will be the key to success, to avoid others stealing the games deserved spotlight.

IV. CONCLUSION

This report has shown that big data can bring lots of benefits to an SME. There are opportunities to help build better levels, choose the right gaming platform, target gamer’s products and also help build a stronger online infrastructure, all from collecting and analyzing valuable data which is relevant to the game. The challenges cannot be ignored though and need to be taken into serious consideration before entering the uncertain world of big data. Finally if big data is used as an aid and not a replacement, alongside a creative game, an SME or any company can increase the success level of their game, and enjoy the rewards it will bring.

V. REFERENCES


Big Data And Recruitment
Hire The Best Qualified Person For The Job, Faster And More Cost-Effectively

Vasiliki Kozi
University of Derby
Derby, Derbyshire
100332523@unimail.derby.ac.uk

Abstract— Big Data has become a useful recruitment tool in recent years. This paper examines how data can help SME’s finding the best qualified person for the job, faster and more cost-effectively. It will count the benefits Big Data brings in recruitment and also it will point some of the problems a business can face by using Big Data in recruitment.

I. INTRODUCTION

A simple definition could be used to describe Big Data is: “Managing, analysing and mining data, no matter how large it is, what format it has and how fast is produced.” It is easy for someone to realize that big data are not about the size, the type or the kind of data, but a number of processes that require computer skills, statistics and business management.

Big Data are characterized by 3Vs: Volume, Variety, and Velocity. It is known that there is a big amount of data available on the internet today (Volume). “There were 5 EXabytes of information created between the dawns of civilization to 2003, but that much information is now created every 2 days.” (Schmidt, 2013) Furthermore data come in various forms, such as relational data, image, audio, video (Variety). Finally, the output data are growing really quickly (e.g. tweets, financial streams) and the inferences must be done in real time (Velocity).

One of the most difficult processes for a business is recruiting. Finding and hiring the right person for the right position is a challenge for every company. In fact 50% of businesses in 2011 admitted that hired a candidate who was not perfect for the job, because finding the right person is an expensive and time consuming process. These days, Big Data is helping a lot of companies to find the best possible talent. Internet has a large amount of information available on potential candidates but wading through all that data it is going to be a time consuming and expensive process. Finding a lot of applicant tracking systems and workforce analytics which are designed for recruiting purposes is very easy today but this is not always the key to success. The point is understanding the importance of each data point and interprets it. Actually nowadays, a lot of people have access to big data with Facebook, LinkedIn, Google+ and other social media. However, the real challenge is to find the right tools in order to analyse it properly.

However Big Data has disadvantages too. A lot of people are claiming that Big Data is invading in their privacy and using their personal information without permission. For example, a lot of companies are buying data that concern candidates’ searches via search engines because they want to learn more things about their interests before they hire them. Furthermore, big data contributes to the elimination of Face-to-Face contact between the employer and the candidates. For this reason there is lack of personal interaction and also companies cannot reach candidates who are not on the web.

II. DATA MINING

A. How Much Data Is Enough?

This question made a lot of companies wonder if they can analyse these huge amounts of information. Before big data businesses had to deal with a lot of interviews in order to find the person they were looking for. That was a problem because they were losing precious time and money. These days Big Data is making hirer’s life easier because now a hirer only has to interview people who are best qualified for the job than the people they used to interview. However, these days there is the problem of huge number of data. For example, if companies used to interview 100 people who were suitable for the job position now they have to interview 100 or more people who are great for the job position. This is making the choice harder and eventually they are losing the same amount of time.

So, how much data is enough?

Unfortunately, there is no simple answer to this question. The answer depends on the particular algorithms employed, the complexity of data, and the relative frequency of possible outcomes. Statisticians have spent years developing tests for determining the smallest model set that can be used to produce a model. Machine learning researchers have spent much time and energy devising ways to let parts of the training set be reused for validation and test.

In any case, where data is scarce, data mining is not only less effective, it is less likely to be useful. Data mining is most useful when the sheer volume of data obscures patterns that might be detectable in smaller databases. (Nanobukva, 2014) Michael J.A. Berry and Gordon S. Linoff (2014) recommend to use so much data that the questions about constitutes an adequate sample size simply do not arise and generally starting with tens of thousands if not millions of pre classified records so that the training, validation, and test sets each contain many thousands of records.
In big data less is more, in data mining more is less but with some caveats. The first caveat has to do with relationship between the size of the model set and its density. Density refers to the prevalence of the outcome of interests. Often the target variable represents something relatively rare. It is rare for prospects to respond to a direct e-mail offer. It is rare and almost impossible for credit card holders to commit fraud online. (Nanobukva, 2014)

The second caveat has to do with big data manager’s time. Since integration of data could be an iterative process, the time spent waiting for results can become very large if each run of a modelling routine takes hours instead of minutes. (Nanobukva, 2014)

III. MATCHING PEOPLE AND JOBS

Organizations want to make the best use of people. In practice, it is rarely possible to match perfectly the requirements of an individual job with the skills and abilities of the people available. Square pegs in round holes are not only bad for the organization; wrongly placed workers are often unhappy and bored, or anxious about being out of their depth. In line with the three basic recruitment strategies, any mismatch between person and job can be resolved in one of the following ways (Drenth and Algera, 1987):

• Select the best qualified person for the job.
• Change job characteristics to fit the abilities of the people employed.
• Train people to perform more effectively (Price, 2007)

Big Data came to change everything in Human Resources. The “right” person approach is easier with Big Data and organization do not have to use the second approach of fitting the person into the organization nor the third, more demanding, approach of recruiting “flexible employees”. They simply find the “perfect candidate” by using Big Data.

The “right person” approach attempts to be objective. It requires clear answers to questions such as:

• Is there a job to fill?
• If so, what tasks and responsibilities are involved?
• What qualities, skills or experience are required to perform the tasks?
• What process will best identify these criteria? (Price, 2007)

The “right person” approach is entirely concerned with the individual, whereas the cultural fit model is consistent with a focus on team working. In practice, the models are easily confused with each other and many sectors apply a mixture of both. Frequently selectors believe they are using the best practice to find the person who meets the specified criteria. In fact, the person chosen is the one whose face fits. All too often resourcing emphasizes the selection of people who fit existing culture and practice at the expense of future needs. (Price, 2007)

IV. COST OF RECRUITMENT

Employee resourcing involves risk and uncertainty. Above all, assessors want to avoid the consequences of picking the wrong person. This may be for the valid reason that an unsuitable person will not perform to required standards. However, selectors are also aware of the consequences of an unfortunate choice rebounding directly on themselves (and their reputations). This encourages selectors to take safe decisions minimizing risk of error. (Price, 2007)

Big Data in recruitment has two main goals. The first one is to hire the best person for the right position and the second one is to eliminate the time, the cost of the process and the risk of errors mentioned above. Recruitment is a very important stage for a company because a bad hire can cost a lot of money and time. In fact, 41% of companies estimate that a bad hire cost them more than $25000 while one in four said it cost them over $50000. Also 50% of the employees hired someone who was not perfect for the job because finding the “perfect” candidate is an expensive and time consuming process.

Furthermore, according to a study from Development Dimensions International (DDI) the survey of 1515 hiring managers showed that over a third of respondents said they made a bad hiring decision because of pressure to fill the position. ‘The cost of a bad hire is much higher than the cost of leaving the role open for a few more weeks’, according to Erker (2007). (Hurst, 2014)

A bad hire has other serious impacts too. A person who is not right for the job can affect company’s productivity and also disorientate the whole team. One employee can cause disarray into an entire department. Companies also run the risk of losing their good employees because they may need to work harder in order to fix the problems that were caused by the bad hire.

Selecting the best qualified person for the job is another problem that companies are facing. In order to find the most suitable person for the job businesses are interviewing a lot of people. Interviews are not the only expensive process companies have to face during recruiting: print or radio advertising, telephone, subscriptions to social media and hiring recruiters or agencies to help are few examples of recruitment expenses.

Big Data can help companies to eliminate these problems. First of all big data will help companies separate the people who are not suitable for the job from the people who actually are. Secondly, big data can highlight the relevant qualities of each person and assign them to company’s requirements. Thereby the interview process will be sorter, less expensive and more qualitative. Finally, by hiring the right person the problem of the bad hire will disappear.

V. SOCIAL MEDIAS, EMPLOYMENT SITES AND BIG DATA

Social Medias have become a mainstream recruitment medium in recent years. It has become normal for jobseekers to scan employment sites and Social Medias on the web for opportunities; so much that many organizations block access to job sites from their workstations to prevent employees from job hunting during working hours. (Price, 2007)

Most large organizations and many smaller ones, make extensive use of corporate websites in their recruitment programs. Typically, general career information is presented in an engaging manner to promote the employer brand and gain interest from prospective applicants. It has become common for
early stages of the selection process to be made accessible online, allowing resumes and CVs to be uploaded, application forms to be completed and preselection tests to be conducted. (Price, 2007)

Social Medias and employment sites are the having a lot of information that a business needs in order to find the right candidate. Using analytics techniques such as text analytics, predictive analytics, data mining and statistics, businesses can analyze data from these sources and gain new insights resulting in significantly better and faster decisions.

VI. METHODS AND TOOLS

The Data era is something new for the recruiting field. For that reason Big Data Analytics shows a lot of promise.

In order to gather all the information they need, businesses are using metrics relevant for the company or the job opening. Visually pleasing dashboards represent the chosen metrics. For example, a framework includes data point that humans cannot manually process. Also are establishing statistical significance among metrics and revealing correlations and insights from data. Finally are presenting further evaluation metrics because sometimes recruiters overlook some of them.

Human Resources Information System (HRIS), A-Check, Electronic Data Capture (EDC), Applicant Tracking System (ATS), Key Performance Indicators (KPI) recruiting informatics and generally analytics platforms are some of these tools.

A. Human Resources Information System (HRIS)

HRIS is a software or online solution for the data entry, data tracking, and data information needs of the Human Resources, payroll, management, and accounting functions within a business. (Heathfield, 2014)

B. A-Check

Integrating A-Check’s background check services into your organization’s HRIS/ATS/VMS system allows end users to operate from a familiar environment, reduces duplicative data entry tasks, decreases errors and speeds the onboarding process. (A-checkamerica.com, 2014)

C. EDC

Modern electronic data capture software is typically web-based and utilizes a thin client. Web-based means that the software runs entirely on a Web server (e.g. Google.com), and thin client means that the only tool you need is an ordinary web browser (without any cumbersome plug-ins) connected to the internet in order to access and utilize the EDC software. (brainstrom.org, 2014)

D. Applicant Tracking System (ATS)

ATS is an effective way to post positions, organize candidates and streamline the hiring process. For many HR professionals, ATSs are typically their first experience with new talent management solutions. (Stupak, 2009)

E. Key Performance Indicators (KPI)

Key Performance Indicators are quantifiable measurements, agreed to beforehand, that reflect the critical success factors of an organization. They will differ depending on the organization. (Reh, 2012)

VII. HUMAN ELEMENT

Big Data is taking recruitment to a new level but also creates a lot of questions about the fact that Big Data in Human Resources is abandoning the human element. Big Data is great at finding the candidates a business needs but it struggles to judge their personalities. Therefore there is a need to humanize big data.

Furthermore, using big data can lead to problems such as less face to face contact. Organizations need to know that their employees are fitting in their culture said Raymond A. Noe, Robert at the Ohio State University’s Fisher College of Business. Raymond Noe argues that a clearly denied culture can help with the employee commitment allow the employees to understand what the company stands for, and provide growth opportunities for those who match well. It follows that enabling potential recruits to understand a company’s culture should be a key part of its efforts to attract and retain high performers. (hrmguide.net, 2014) This is a good opportunity for employers to ask themselves if Big Data is always the best way to recruit.

Another problem that businesses are not taking into consideration is that some jobseekers are not on the Web so they cannot reach them. This kind of problem is actually decreases continuously because nowadays almost everyone has access to the web.

There is also the problem of privacy. There is no doubt that using Big Data in recruitment is a beneficial method for a company. However, overusing these metrics and systems could only lead to abuse. A well-informed and conscious company respects the legal framework around data privacy.

Last but not least, with regards to social media some job seekers find that it gives off the wrong image for them. For example, a Facebook account contains personal information that many times do not concern a person’s job abilities and qualifications.

VIII. CONCLUSION

Taking all the above mentioned into consideration it can be said that Big Data are taking Recruitment in a new level. Companies can use data in order to find the perfect candidates faster and more cost-effectively. However they must be aware of the legal framework otherwise they can face serious problems.

Cost was the bigger problem in recruitment. Interviews, CV reviews, advertising, attendance at recruitment shows and salaries of in-house recruiters were some important costs of recruitment.

Social Medias and job websites were helping companies in finding the “right person” but that was not enough. Big Data is not only finding people who have the appropriate knowledge for the job but gathers and aggregates information about them.
More important is interpreting and understanding the importance each data point.

Judging someone by his personality is something big data is forgetting. Abandoning the human element can cause problems to a company. For this reason Big Data must be used as a tool and companies should understand that they cannot lean completely on Big Data for recruiting.

Big Data can take you from reactive to proactive, from gut instinct or habit to evidence-based decision-making, from “I think” to “I know.” (theundercoverrecruiter.com, 2014)

IX. REFERENCES


Big Data, Big Lie

Daniella Kypri
Department of Business, Computing and Law
University of Derby
Derby, United Kingdom
d.kypri1@unimail.derby.ac.uk

Abstract— This report will investigate the negative outcomes of big data, and how the inaccuracy of data leads to companies determining useless and incorrect results. It will also look at how the use of social media can benefit SME’s by providing them with more accurate data and different ways of analyzing this data, which may prove more successful.

Index Terms—Veracity, Inaccuracy, Social Media, Data analytics.

I. INTRODUCTION

When analysing big data, the 3 V’s; volume, variety, and velocity are taken into consideration. It seems that the fourth and most important ‘V’ is forgotten about entirely. “Veracity in the context of Big Data connotes not only to the accuracy of the data but also the reliability of the source, the context, engagement, and interaction that triggered the data to be generated, the methods, transforms, analytics used in the information extraction, and actual information derived from it.” (A. Bhambhri, 2013.) Having a large amount of data to analyse is great, but what if the data is useless?

Data has increasingly become more valuable to companies, giving them an insight into who their consumers are and what are their needs. However, it is unclear to which extent the data that they are provided is accurate. According to the Global Research Report of 2014, “The findings from a new Experian report, indicates that this flawed data is wasting both time and resources totaling £197m in the UK alone.” (netimperative, 2014.) Furthermore, the study indicates “three quarters of UK organisations are losing potential revenue due to poor contact data, and 94% believe they have poor data quality in their organisations.” (netimperative, 2014) Big data can be extremely valuable to companies, when it is accurate and analysed correctly. “Big data promises new breakthroughs. But companies are putting too much emphasis on the amount of data they are able to gain. And completely overlooking that the more data they have, the harder it is to analyse. Ultimately meaning that they are coming to incorrect conclusions.” (L. Tucci, 2013.)

II. BIG DATA INACCURACY

It can be so easy for the data that is provided to companies to be wrong. “More than 90% of organisations report at least one type of common error in their contact data, from missing information and inaccuracies, to out-dated and duplicate data.” (netimperative, 2014.) Information is often entered incorrectly; there are various ways that data can be incorrect. For example something as simple as an address, which is something that can constantly change. “Addresses are a major issue for the government. An individual’s address is one of the primary identifiers that the government and its agencies use to identify constituents. Unfortunately, addresses aren’t always permanent and people move.” (M. McLoud, 2013.) People’s circumstances change, they get divorced, they may move houses, and they may become unemployed. But their data stays the same. “However, as more sparse information gets shared across lines of business, the risk of ingesting old or inaccurate data increases, leading to potentially biased or false conclusions.” (A. Bhambhri, 2013.)

Furthermore, as big data is so big, the more data there is, the more likely that false correlations begin to appear. “Because of excess data as compared to real signals, someone looking at history from the vantage point of a library will necessarily find many more spurious relationships than one who sees matters in the making.” (O. Ogas, 2013)

![Spurious Correlations](image)

Fig. 1. Ogas (2013)

Figure 1 indicates that the higher number of variables shows a higher number of correlations. “Falsity also grows faster than information.”

Companies are putting too much emphasis on the amount of data they are able to gain. And completely overlooking that the more data they have, the harder it is to analyse. Ultimately meaning that they are coming to incorrect conclusions. “Despite the enormous growth in the amount of data—and the success of a few companies—the reality is that deeper insights
for most organizations remain elusive. Data analytics is only a tool. When we use it as a strategy, we make assumptions about people and their behavior that have no genuine connection to the real world.” (Rasmussen & Madsbjerg, 2013)

A. Cookies and Browsing Data

Furthermore, using historical browsing data to predict future interests and behaviour could also be extremely incorrect. (A. Avner, 2012.) It has been noted repeatedly that the foundational units of “big data,” the cookie and the look-alike model, are often extraordinarily inaccurate.” (A. Avner, 2012.) Knowing what people are looking at, what they are searching for, what they are posting, is something extremely valuable. “Most individual bits of data are worthless; it's only in the aggregate that they become valuable.” (The law of large numbers, 2013) And when this data builds up, companies may be able to determine something useful. However “people share computers, so you’ll never know at any given moment whether it’s my girlfriend or me that your algorithm has bought — but some of it is just the nature of the platform” (A. Avner, 2012.) More than 2 people live in an average UK household (Office for National Statistics, 2011.) It is likely that they share a computer, and even something as simple as gender could be inaccurate from the data collected from cookies and browsing history. “We’ve seen agencies run tests against the validity of cookies on a data exchange. The gender is wrong 30-35 percent of the time.” (A. Avner, 2012.)

B. Inaccurate Data Analytics

It could be argued that the data is not the issue, but the way it is being analysed. “Big data is not about the data, it's about the analytics, according to Harvard University professor Gary King -- and, boy, are there some really bad analytics out there.” (L. Tucci, 2013.) Even if the data that is provided is completely accurate, analysing data incorrectly could put it all to waste. For example, Gary King talks about a big data project set out to use social media in order to predict the unemployment rate in the US. A common analytics approach was taken, a word count to calculate a set of words that applied to unemployment, such as ‘jobs’ and ‘classifieds’. The aim was to look for a correlation between the amount of these words used on social media per month, and the monthly unemployment rate. However, the study did not go to plan, as Steve Jobs died the same month, and the amount of times the word ‘jobs’ was used on social media skyrocketed. (L. Tucci, 2013.) Something completely unrelated caused the whole project to fail. Big data is complex, and so the analysis must be well planned out, in order to avoid situations like these. There are various approaches to analysing data, but to truly come to a significant result, the analysis must be done in a way that avoid any possible mistakes.

C. Incorrect Conclusions

Companies are focusing on the amount of data they are able to receive, and constantly finding new ways to obtain this ‘valuable’ data. However, the quality of the data that that are obtaining is being overlooked. “Unfortunately, many data analytics initiatives have under-delivered or failed to deliver because there has been little focus on data quality.” (M. McCloud, 2013.) Due to the inaccuracy of data, and the poor methods used to analyse it, companies are coming to incorrect conclusions, deeming the data and process completely useless and unbenefficial. “One of the greatest benefits big data offers is the gift of foresight. By using predictive analytics, an organisation can make sound judgements as to how certain events are likely to unfold. However, relying too heavily on these predictions or taking them at face value could mean you set yourself up for a rather sizeable fall.” (The Big Data Insight Group, 2013.)

III. RISKS OF USING BIG DATA

As a small business, investing in big data projects is a huge risk. “As an example, many customer relationship management (CRM) implementations and data analytics initiatives fail to deliver on promises simply because the data that exists is inaccurate or incomplete.” (netimperative, 2014) It is important for companies to first have a clear understanding of big data and what exactly they are looking to accomplish. In order to make sense of the onslaught of Big Data, organizations need a way to view data as an entire landscape of related information, not just an endless array of unrelated points.” (A. Bhambrri, 2013.) For a small business to invest in big data, they should feel confident that the data is accurate, and that they will be able to efficiently analyse it, in order to come to a useful conclusion. Michael Passingham says, “I like to think it's about 'ability' because data alone contains useful information in a concealed fashion. The ability thing here is the ability to understand and extract the information that's inside the data to make it become useful.” (M. Passingham, 2014) There is in fact potential in big data, the risk for businesses is to have the resources and ability to use it in a way that will be beneficial to them, instead of just using it to say that they have. “People don’t really care if the data is accurate. They care that it’s perceived to be accurate enough that they’ve done their job.” (J. Marshall, 2013)

IV. OPPORTUNITIES

For small and medium enterprises wanting to target market customers is obviously something which can benefit their company, and provide them with a new outlook as to who their consumers are, and how they can better reach them. However, for a small enterprise the cost of collecting big data on their own is not only very expensive, but also extremely risky. The lack of accuracy could cause entire projects to go to waste, which is something that SME’s cannot really afford.

Nevertheless, there are other ways that small and medium enterprises can use “big data” that is less risky and costly. “By using non-traffic big data sources it is possible to increase accuracy.” (C. Orton-Jones, 2013) With the use of social media sites, such as Facebook and Twitter, companies can use already provided big data in order to target market consumers. The risks of this are astoundingly less, as this service is more reliable, with a higher accuracy of answers. “What makes social media different than other data sources is the personal nature of the data created.” (C. Uganec, 2013)
V. SOCIAL MEDIA OPPORTUNITIES

As previously stated, one of the biggest downsides of big data is the inaccuracy; data is old and not kept up to date. Social media is a way to decrease these inaccuracies. Social media users are just that: users. They keep their Facebook profiles and details up to date, and post things that are relevant to their interests. This could be one way to eliminate irrelevant data. As Amit Avner states, “The other solution is a blunt instrument that puts a lot of companies out of business but will absolutely improve data accuracy: the Facebook ad network, where all targeting is based on declared data and there are few look-alike models, only (anonymous) individuals that can be bought on the basis of granular knowledge and not inference.” A. Avner, 2012) This way of using social media data could not only benefit the company, but also the social media users. “Potentially, this data could also make social networks like Facebook and Pinterest do a better job of showing users what they want to see, rather than content and ads they'd rather not waste time on.” (C. Smith, 2014.) Amit Avner also goes on to say that if Facebook “drops an identifier with every login, it solves the computer-sharing problem, the cookie-deletion problem, the mobile-targeting problem and almost all other big data problems in one swoop.” This could potentially eliminate many of the various causes of inaccurate and incorrect data from social media sites. However something of such a large scale, which could prove to be very successful, comes at a large cost. Additionally this does not fully exclude the unpredictability and imprecision of data. “Social Media data flows can be highly unpredictable with periodic peaks. Such data loads from what’s trending in social media, mixed up with unstructured data are even more challenging to manage yet interesting to explore.” (J. Bowden, 2014) Furthermore, “By leveraging social media analytics tools, organizations can not only connect to their customers in a better way, but can also get unique insights into consumer preferences, market trends, brand awareness, customer buying behavior and spending patterns.” (J. Kohli, 2013) By using social media as a way to collect and analyse data, companies have a better chance of really getting to know their consumers, and then tending to their needs. This is ultimately the point of big data, to find new ways to attract customers.

A. Social Media Analytics

Small and medium enterprises should focus on one resource to gain data. Social media has become one of the fastest ways to generate data. “There are over a billion active users of social media network worldwide, many of whom are frequently active and can be connected by means of their smartphones and tablets. Social media indeed has become a main communication network in the daily lives of people around the world.” (J. Bowden, 2014) Social media is a way for companies to gain access to relevant data, which is more up to date than any other form, eliminating the chance of data being out of date. Therefore, companies should be investing in this form of data, if any. “In fact, social media now embodies the leading and biggest source of consumer data. Just imagine the hundreds of thousands of posts about a company’s products or services that have been published every day!” (J. Bowden, 2014) Furthermore, there are also various efficient ways to analyse data provided from social networking sites. “There are a number of types of software tools for analyzing unstructured data found in tweets and Facebook posts. In addition to to text analysis, many enterprise-level social media tools will harvest and store the data.” (M. Rouse, 2012) The results of data which is already provided by social networking sites, is much more likely to be more accurate, and based on what users are saying in ‘real time.’

B. Facebook Ad Network

Data is now everywhere, but much of it is nearly worthless. (J. Marshall, 2013) Small and medium enterprises need a reliable way to make use of the data that is available to them. They need a strategy that is reliable and efficient. “Facebook Inc. FB +2.92% is planning a mobile-ad network that will allow the company to tap its vast reservoir of data about users to help marketers target ads on other services, according to people familiar with the matter.” (Albergotti & Marshall, 2014) This technology could potentially change the way that companies use data. However, there is still the issue of how this data will be analysed. “Facebook’s new ad network won’t increase the amount of data the company gathers on its users, but it could increase the amount of targeted ads Facebook users see across the mobile Internet. Users won’t be able to opt-out of receiving the ads.” (Albergotti & Marshall, 2014) The ad network could be a great success to companies, due to Facebook’s very active user base.

Although this technology will not mean more data, this could be a good thing. The ad network could provide a way for companies to reach out to a large range of users, who are interested in what they have to sell. Instead of spending time and resources on analysing data which may prove incorrect and irrelevant.

VI. BENEFITS

There are various benefits of taking this approach of data collection and analysis. Social media provides real time data, which is both accurate and relevant. “Traditional data storage systems were not designed for real-time analysis but new technologies can now provide live information and data analysis can be accomplished in real-time.” (Parliament UK, 2014) Social media analytics also provides data that is relevant to a company, so that they can use it to target customers who would be interested in specific products of services. “By leveraging social media analytics tools, organizations can not only connect to their customers in a better way, but can also get unique insights into consumer preferences, market trends, brand awareness, customer buying behavior and spending patterns” (J. Kohli, 2013) Furthermore it gives businesses an insight into what customers are saying about them, negative and positive comments which could help businesses adjust.

VII. CONCLUSION

Big data may seem like a great opportunity, and there has been proof of great outcome from the use of big data. However, for a small enterprise the risk of using big data is simply too
high. Big data lacks accuracy, and without the correct resources, all of the data could possibly go to waste, or result in incorrect conclusions that will not benefit the company. “Big data is only valuable if it tells a story. The fuller the story your data tells, the better you’ll be able to take advantage of that data. While recognizing a trend can help you make better decisions, understanding the cause behind that trend is even more valuable.” (C. Uganec) Accuracy is key when analysing data, and if the data that is collected is out of date or incorrect it is of no use. “Big data may mean more information, but it also means more false information.” (O. Ogas, 2013) For small and medium enterprises there are other options. The process of collecting big data from the beginning is just not worthwhile. But using already provided data, from social networking sites, may prove more successful, accurate and have greater results. “Then we won’t be talking about Big Data; we’ll be talking about Good Data.” (A. Avner, 2012.)

VIII. REFERENCES


Big Data Opportunities
For Small And Medium-Sized Companies

Evaldas Luksys
University of Derby
Derby, United Kingdom
E.Luksys1@unimail.derby.ac.uk

Abstract—This report mainly focuses on one domain: big data opportunities for small and medium-sized companies. The first part of the report explores a domain of big data, identifies opportunities of big data in a variety of sectors, including public and private, and presents real business cases to prove the usefulness of the big data technologies. The second part of the report focuses more on SMEs by exploring various options of how big data and its technologies could be obtained and what opportunities the big data technology could offer for small and medium-sized companies. Also the report explores and identifies the key factors that should be considered when implementing the big data technology in SMEs, improving their businesses from various aspects and competing with their rivals with a help of big data. By covering these particular topics this report aims to provide SMEs with valuable information that could be utilized to improve various SMEs businesses from various aspects and thus increase their competitiveness in a global market.

Index Terms—Big data, Big data technologies, SMEs, Big data and SMEs, Big data opportunities.

I. INTRODUCTION

The big data is a commonly met term that describes the information which cannot be analyzed by utilizing traditional data analysis methods and tools (Zikopoulos et. al., 2012) and is found in both: structured and unstructured forms (Ammu and Irfanuddin, 2013). To grasp the significance of big data Gartner (2011) has announced the big data being rapidly growing worldwide which by estimates could be increasing by at least 59% annually in volume. Today’s organizations, including the main subject of report – SMEs, have to embrace big data and its challenges in order to leverage the opportunities (Greengard, 2013) and benefits offered by big data (McGuire, Manyika and Chui, 2012). The next section briefly introduces what SMEs are and the situation of big data technology in this particular type of organizations.

II. SMES AND BIG DATA

SMEs abbreviation stands for small and medium-sized enterprises which are determined according to two key factors: the number of employees, a total of balance sheet or turnover. According to this EU definition, companies that have the number of employees below 250 and turnover (equals or less than €50 million or equals or less than €43 million of balance sheet total) are held as SMEs (European Commission, n.d.). The significance of this particular topic is that SMEs is the most met type of a company worldwide. For example, in United Kingdom “99.9 per cent of private sector businesses are SMEs, employing an estimated 14.4 million people, 59.3 per cent of private sector employment. Their estimated combined annual turnover of £1,600 billion accounted for 48.1 per cent of private sector turnover” (Department for Business Innovation & Skills, 2013). In the European Union SMEs represent 99% (Department for Business Innovation & Skills, 2013) and in United States 99.7% of enterprises (SBA, 2012). By seeing how big SMEs market share is there is a need to explore and present opportunities and benefits of big data. There is a need to prove that big data can be useful and thus encourage small companies to take action and leverage big data’s potential, since according to SAS (2013) report, only a small fraction of SMEs incorporates big data technologies which is thought to be less than 0.2% of all SMEs.

III. BIG DATA DEFINITION

When it comes to defining big data, a variety of sources have gathered a substantial number of defining terms which by big data community are called big V’s (Grimes, 2013). For example, IBM (n.d.) has identified four V’s including Volume, Velocity, Variety and Veracity. According to Gartner (2011) there are only 3 V’s including Volume, Variety and Velocity. Other source (Katal, Wazid and Goudar, 2013) along with the above mentioned V’s presented more big data characteristics including Variability, Complexity and Value. Clearly each and every attribute provides with a new perspective of looking at big data and the following section analyses these properties to gain deeper insight of the opportunities that are offered by the big data and its technologies. The following section will cover some of big data V’s including Volume, Velocity and Variety.

A. Opportunities according to V’s

1) Volume

An increase in volume according to EMC (2012) in 2020 will result in 40,000 exabytes of data compared to 2005 with only 130 exabytes worldwide (which indicates that big data volume doubles in less than two years worldwide). Having the right technologies to cope with such a growth of data volumes the opportunity behind this characteristic exists – the more data used, the better insight could be produced for the chosen business marketplace and customers (Zikopoulos et. al., 2012).

2) Velocity

Big data velocity refers to the speed of data creation. Real-time or nearly real-time data provides businesses with an
opportunity to analyze the streaming data in order to compete with rivals more efficiently (McAfee and Brynjolfsson, 2012) and understand the latest trends and needs of the customers. To understand how big the impact of velocity can be IBM (n.d.) has provided with some of the projected estimates for the future: in 2016 there will be around 18.9 billion of network connections.

3) Variety
The variety provides an opportunity to gather data from various sources such as text messages, images or texts posted to social networks, signals of GPS (cell phones), data from a variety of sensors and etc. (McAfee and Brynjolfsson, 2012). This means that all the types of data are being captured including structured, semi-structured and unstructured (DataStax, 2013) which provides organizations with more complete data to utilize.

IV. OPPORTUNITIES OF BIG DATA

Big data is taken seriously by various organizations and is an important subject to be incorporated into companies (Gartner, 2013) for its broad and remarkable potential and benefits (NIST Big Data, n.d.). According to Market and Markets press release (n.d.) “the Global big data market is estimated to be $14.87 billion in 2013 and expected to grow to $46.34 billion” by 2018. Gartner (2013) revealed the results of the survey which stated that 64% of organizations had already invested or were planning to invest in a technology of big data during 2013. This part of the report focuses on some organizations that noticed a huge potential of the big data technology in a variety of domains and provides brief excerpts of the situations where big data’s opportunities were revealed.

A. McKinsey Global Institute
McKinsey Global Institute conducted a research and analysis on the impact of big data in 5 different domains including “healthcare in the United States, the public sector in Europe, retail in the United States, and manufacturing and personal-location data globally” (Manyika, et.al., 2011) and concluded that big data could have a huge impact and thus add value to every sector. For example, retailers by utilizing big data could increase their operating margins over 60% and “users of services enabled by personal-location data could capture $600 billion in consumer surplus” (Manyika, et.al., 2011).

B. Tesco and IBM
Tesco has been working closely with IBM to achieve some of the goals during the specific time frames. By using big data Tesco aims to “reduce its carbon footprint in Ireland by 50% by 2020” (Goodwin, 2013) which also means the savings of energy from 5% to 10% annually. In addition, Tesco targets to save electricity costs by optimizing temperatures of its refrigerators across 3000 stores in Ireland and United Kingdom which by predictions should save up 20 million euros annually.

V. REAL BUSINESS CASES

To prove the potential and usefulness of big data and its technologies the report introduces the readers to a couple of examples of real businesses where big data technologies were utilized to achieve substantial results.

A. Time Reduction
Macy’s company with a help of big data technologies was able “to reduce the time to optimize pricing of its 73 million items for sale” (Davenport and Dyché, 2013) which originally took 27 hours and was reduced to only 1 hour. In another example (IBM, 2013), Mizrahi Tefahot Bank was able to reduce time to market new financial offerings by 99%. Also according to IBM (2013) report the company called Qualcomm reduced time to market from several months to a couple of days.

B. Cost Reduction
According to SAS (Davenport and Dyché, 2013) big data has helped significantly to save large amounts of money from the driving costs. By reducing 85 million miles of daily routes the big data led to savings of 8.4 million gallons of fuels in 2011. UPS estimated that saving “one daily mile driven per driver saves the company $30 million” (Davenport and Dyché, 2013). The Second example (IBM, 2012) is an agriculture business that has gained many benefits from big data technologies including the reduction of harvesting costs by 5%.

VI. SUMMARIZING THE KEY OPPORTUNITIES
The opportunities and benefits that were previously mentioned have revealed how big data is impacting various organizations, however there was mentioned just a fraction of opportunities and benefits which big data provides. According to Actuate (2012) report, there are many opportunities and all of them can be divided into separate sections. For example, report identifies 4 major areas of business where big data provides huge opportunities. These areas include the improvement of operational efficiencies and revenue generation, enhancement of customer experience and identification of beneficial business strategies.

VII. AVAILABILITY OF BIG DATA TECHNOLOGY

The previous part mentioned the increasingly growing availability of big data technologies worldwide. Finch (2013) complements this idea by stating that the big data technologies are more accessible for wider audience than they were before. The report in this section aims to find out what influences this trend of growing availability of big data technology and to which extent this availability exists for businesses including small and medium-sized organizations.

A. The Availability Of Big Data
It was previously stated that volumes of big data are growing at immense pace and by 2020 there will be 300 times more of data worldwide than we had in 2005 (Zikopoulos et. al., 2012). These changes are influenced by one major factor – big data comes from almost every imaginable field (CSC, n.d.) and from various data sources including transactional data (including invoices, payment orders, records of storage, delivery records), machine data (including industrial equipment, real-time data from variety of sensors and web
logs), social data (including social media services such as Facebook, Twitter, YouTube and etc.) (Oneopintiq, n.d.). The key opportunity of big data’s availability lies in an increase of such sources that could be utilized for data gathering purposes. The predictions (Gartner, 2013) for the growth of internet of things (that excludes personal computers, tablets and smartphones) suggest an increase from 0.9 billion units in 2009 to 26 billion units in 2020. ABI research (2013) has stated that by 2020 there could be 30 billion wirelessly connected devices compared to the year 2013 with only 10 billion devices. In case the SMEs struggle to gather data from mentioned sources, the research conducted by University of Harvard (Simon, 2013) proposes a viable solution: “Even a lack of data isn’t an insurmountable obstacle to harnessing analytics”. In addition, Big Data Startup (n.d.) has provided with over 17 big data providers where customers can buy or even download data (from government and organizations) for no cost. There are even more options to obtain data, for example, “Nike for example shares data from all its suppliers with the rest of the industry. This allows other organizations in the supply chain also to populate and use the database and make better decisions” (Rijmenam, 2013).

B. The Availability Of Big Data Tools

By analyzing big data tools’ assortment, there were found various paid and open source solutions. Since small businesses usually cannot afford to spend significant amounts of money on big data tools, there are lots of open source software tools offered, for example, according to Big Data Startups (n.d.) “The Big Data Open Source Tools landscape is growing rapidly” and provided a list of more than 120 open source tools that could be used to utilize big data. In case SMEs do not know how to use big data tools, the enterprises can rely on other companies to deal with their big data (Simon, 2013). In case the business can afford more advanced big data tools, there is also a variety of big data solutions offered by companies such as Oracle, SAS, Microsoft, IBM and others (Henschen, 2014). One article (Simons, 2013) has concluded an idea of big data tools and SMEs: “embracing data tools doesn’t have to mean a huge financial commitment for those seeking to affect positive change on their businesses”. The same opinion was stated in Forbes article (Thusoo, 2013) where technologies of big data (including both: data and tools) are said to be a significant part of SMEs in near future since commoditized big data (including operations and management) will offer inexpensive and risk-free solutions.

VIII. Big Data Opportunities For SMEs

Big data technology has been popular for a while among large corporations and was considered to be too difficult and too expensive to be utilized by small companies (Devaney and Stein, 2013). However, at these days the situation is changing and more small companies start gaining the benefits offered by big data due to increasing “availability of data and analytics” (INTUIT, 2012). The same report (INTUIT, 2012) states that big data tools and sophisticated analytics are soon to move from professionals to every day users, big data large volumes are becoming more available than ever before and the big data revolution promises a broad range of opportunities and benefits. In addition, the same report points out the fact that thousands of businesses at initial stages are formed and developed further with a significant help of big data. One of the examples of small business that uses big data was presented in Forbes article (Devaney and Stein, 2013). The Spillers Group that owns three restaurants gathers data from every restaurant and by combining it the company is able to optimize some of the processes, more accurately divide payments to their employees according to their performance and save up to 10% of labor costs which saves Spillers thousands of dollars. Another SME – Jetpac (Simons, 2013) was able to produce a new automated tool to help their customers to find the best pictures that could be used within Jetpac’s system. The solution was found in 3 weeks for $5,000 using Kaggle platform which saved a lot of money and time for Jetpac.

IX. SMEs - Competing With Big Companies

As stated in previous sections big data can offer a variety of benefits for organizations, however SMEs can find it difficult to compete with big companies in big data war for various reasons (Passingham, 2013). This section briefly explores how SMEs can tackle this particular issue and utilize big data more effectively in a cost-effective manner.

Since not every large company is able to utilize big data and the big data technologies effectively (Shead and Techworld, 2014), majority of SMEs with considerably less or no available funds at all are not able to take an advantage of big data. For this reason, SMEs should utilize the big data technologies wisely and deeply focus on a few aspects of business “such as sales in a specific sector, or performance metrics during peak versus low seasons” (MacInnes, 2013). In one article it was pointed out that “bigger is not necessarily better” (MacInnes, 2013) which implies that SMEs do not necessarily have to focus on collecting and analyzing vast amounts of data since it could become too complex to utilize big data effectively (Puri, 2014) and utilization of smaller datasets could also help to improve the business (MacInnes, 2013). To lower the costs for collecting the data, implementing and utilizing the big data technologies SMEs can always use options that were previously discussed in this report which include buying already collected data, obtaining publicly available datasets and using free big data tools. Also low cost solutions could be utilized such as CRM technology that enables tracking of the customers, gathering of data and production of visuals that help to determine further ways to improve various processes and strategies of the businesses (Salesforce, n.d.). Also, in case the SMEs have the data, however do not know how to utilize it, there exist various projects, for example, Kaggle which enables the companies to get their data explored by competing teams for the money prize which is given to the most successful team (Simon, 2013).

From the paragraph above it is clear that SMEs have plenty of options to consider. By combining various techniques, having a clear strategy, investing some money and time, SMEs
could significantly improve their businesses from a variety of aspects (Rijmenam, 2013) and thus compete more aggressively and effectively with their rivals.

X. CONCLUSIONS

In conclusion, the report has analyzed and presented big data benefits, opportunities and gave much of the focus regarding large corporations, SMEs, integration and the availability of big data and its tools. The report has shown how vast and prospect the big data technology is, the ways it could be incorporated into the current business environments - which should encourage more SMEs to start doing what big corporative businesses have been doing for years. The provided knowledge should enable SMEs to understand big data and the big data technologies from a broad perspective and see these technologies as an additional tool to augment and improve their businesses. Also the provided knowledge should make the businesses more aware of the options they can have which could decrease the effort and money to incorporate and utilize big data. Also, since the report covered big data opportunities for SMEs topic broadly, there is a need to carry out more in depth research to provide more information about explored topics within this report to further educate SMEs about big data and big data technologies.

XI. REFERENCES


Zikopoulos et al. (2012) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. US: IBM.
The Opportunities And Challenges Of SMEs Using Big Data Within The Travel And Hospitality Industry

Mawhinney, Liam
Faculty of Business, Computing and Law
University of Derby
Derby, United Kingdom
1.mawhinney1@unimail.derby.ac.uk

Abstract—The adoption of Big Data Analytics by SMEs is extremely low. The travel and hospitality industry contributes largely towards a vast amount of data being produced daily, and small hotel owners have an opportunity to gain a competitive advantage over competitors by using Big Data to improve their forecasting accuracy, marketing effectiveness and customer experience. However the challenges relating to the uptake of Big Data analytics must first be acknowledged and overcome in order to be successful.

Index Terms—big data, hospitality, travel, hotel, SMEs

I. INTRODUCTION

The travel and hospitality industry is highly competitive, with an estimated 750,000 hotels established worldwide (TripAdvisor, 2013). SMEs within this industry must therefore identify and secure a competitive advantage over the larger, more sustainable hotels to stand a chance at succeeding. This competitive advantage could be the effective use of Big Data analytics. Research indicates that over 71% of all organizations are yet to implement a Big Data strategy, therefore implementing an effective strategy should become a priority of all SMEs within the industry (Rjmenam, 2014).

Big Data can be defined as large amounts of data, gathered from a variety of sources and used to identify meaningful trends and relationships (Safran, 2013; Steinberg, 2013). This may seem intimidating, however due to the customer/service-driven nature of the industry, it is likely that most hoteliers (hotel owners) have already gathered vast amounts of data but do not have the skills, technology or knowledge to analyse and use the data effectively (Vertolli and Fitzgerald, 2013; EyeforTravel, 2013).

When analyzed correctly, Big Data could reveal trends in current and past business activity, customer behavior patterns and help manage revenue and marketing strategies (Davenport, 2013). However it is also important to be aware of the challenges and risks associated with Big Data analytics, to ensure that this competitive advantage leads to success and not failure.

II. GATHERING BIG DATA

The travel and hospitality industry is constantly gathering and creating data in many forms from customer surveys, website statistics, photographs and social media statuses uploaded by customers, third-party and supplier provided data to in-house system generated data (Vertolli and Fitzgerald, 2013; EyeforTravel, 2013; Adshead, 2013). Viewing this data collectively is the beginning of Big Data analytics.

A. Customer-generated Data

What is truly unique to the travel and hospitality industry is the vast amount of data uploaded by customers without request from hoteliers. For example, the traditional booking process now involves the use of customer reviews, with over 81% of travellers reviewing them as necessary prior to booking (Bartnick, 2014). These reviews are normally generated by customers through third-party sites such as Trivago.co.uk and Tripadvisor.co.uk, and are mostly uploaded without request (Bartnick, 2014).

Increased social media use has also had a significant impact on the amount of data available. The majority of travellers will upload photographs to Instagram/Facebook/Twitter, which can then be used by hoteliers and potential customers to identify satisfaction levels, key attractions and information about the author of the photograph (Vertolli and Fitzgerald, 2013; Redzy, 2014). A new mobile application, Jetpac, goes one step further and uses object-detecting software to extract useful information from photographs to provide in-depth recommendations of local attractions (such as hotels); using details such as a high levels of smiles and men with moustaches in the pictures (Warden, 2013). This level of analysis would provide unique observations and relationships otherwise unnoticed, and the data is all provided by the customer without costly traditional approaches (Smith, 2014).

B. System-generated Data

It is also likely that various in-house and external systems, will have collected vast amounts of data which can be used for historic analysis. For example, customer information from booking information, website statistics, restaurant and bar purchases and management information from past promotion efforts, occupancy records and financial records (Vertolli and Fitzgerald, 2013; Innroad.com, 2013). The data collected from these systems, once analysed and viewed collectively, will reveal trends in past approaches to management behaviours such as advertising and promotional campaign effectiveness, and provide information about past customers for future use.
III. OPPORTUNITIES AND BENEFITS

It is important that the Big Data gathered from the above is used effectively to produce interesting and valuable results. With over 69% of companies using Big Data analytics spending their time and resources on preparing the data and various systems, it is essential that these end results and benefits remain insight to avoid missing an opportunity for competitive advantage (Ventana Research, 2012). It is important that preparation occurs (ensuring that the correct systems, security measures and technology are in place), however results must be obtained to justify these efforts. Three potential benefits and opportunities are explored below, relating to the three main interests of SME hoteliers: financial benefits, marketing/managing benefits and customer benefits.

A. Forecasting Demand and Reducing Costs

There are traditionally two approaches to forecasting demand (occupancy levels): using previously stored data to forecast similar future demand (for example, using last year’s occupancy level data from April 2013, to predict occupancy levels for April 2014) and using current data from various sources (for example, weather forecasts, current and past occupancy levels, forthcoming events, social media trends) to predict potential demand (Hotelogix.com, 2014; Uttamchandani, 2013; Davenport, 2013). Big Data analytics has the potential to use both approaches, combining previously stored data with real-time data which changes with current trends and actual statistics to accurately predict demand. In a fully established system, this information could be sent on a daily basis to all relevant personnel (marketing, finances and hoteliers), allowing them to make more informed and effective decisions (Davenport, 2013). For example, by accurately predicting demand for the next year, promotions and marketing campaigns can be increased and decreased as required.

The benefit of using this approach is that should a situation change, in the market or organization, the forecast will also change to take this into account; allowing for more flexible management of business activities such as marketing and budgets (Coleman, 2014). This also means that the management of expenses can become more flexible. For example, if an increase in occupancy levels is accurately noted, additional housekeeping and catering staff (and their costly resources) can be requested to suit demand (Uttamchandani, 2013).

B. Targeted Marketing

Big Data analytics allows for more informed and fact-driven decisions to be made throughout the organization. As success is usually determined by occupancy levels, keeping them high is a priority for all hoteliers and management personnel; this requires effective and cost-efficient marketing.

1) Targeting Specific Customers

Effective marketing means targeting the right customer, with the right promotion, at the right time (EyeForTravel, 2014). This traditionally requires specific research carried out with the intention of categorizing customers into specific groups to target promotions. For example, are the majority of customer’s business commuters or groups of young people? (Miller, 2013; Salas, 2013).

However using analytics, the focus is taken away from identifying groups of customers, and placed on identifying the characteristics of specific, high-value customers (Hotelogix.com, 2014; Davenport, 2013). These are the customers who spend the most, whether that is through additional costs (such as purchasing high amounts of food and beverages from the bar) or through frequent visits.

By identifying these customers and identifying trends in their behaviour, future promotions specific to their needs can be created and sent at the most influential and effective time (EyeForTravel, 2014). For example, a business commuter may visit on a monthly basis and order all meals from the in-house bar/restaurant, an offer sent via email a week before their next visit, with coupons for their favourite wines and meals, will be satisfy their specific needs and lead to increased satisfaction and better reviews (Greengard, 2013).

2) Targeting Online Customers

As mentioned above social media is becoming extremely popular with travellers, with around 52% of all travellers sharing their experience on Facebook alone (Redzy, 2014). This also relates to an increase in online bookings, with over 57% of all reservations made online (Bartnick, 2014). Not only does this lead to an increase in the availability of valuable data, but it also means that online advertising can be used more successfully. This is a much more effective marketing strategy, and results in lower costs (compared to traditional printing and delivery methods) and higher customer conversion rates.

More personalized advertisements can be sent to customers preferred booking method (whether a smartphone, tablet or PC), and interesting trends can usually be identified between customer behaviours and their chosen method. For example, 65% of same-day reservations are made using smartphones (Bartnick, 2014). Therefore advertising through smartphone applications should reflect immediate offers such as free drinks or reduced prices.

C. Personalization: Improving the Customer Experience

All hoteliers understand the need for demand and high occupancy to succeed; this is achieved through first reaching out to new customers and then maintaining (and hopefully exceeding) their expectations to retain their frequent custom. The service provided by hotels must therefore be excellent; Big Data analytics can help with this too, by providing the hotel with the means of personalizing their services to each individual customer.

Using the same methods as explained above, hoteliers can identify the demands of each individual customer prior to their stay, and prepare their services to meet these demands. For example, if a family reserve a room and photographs on their Facebook and Instagram profiles show that they enjoy outgoing activities, leaflets on local attractions could be placed in their room prior to their arrival and their favourite meals (identified through their previous orders) ordered in preparation for their stay (Rijennam, 2014). By making each service/department aware of certain customer behaviours and trends (for example,
marketing, finances, catering), a customers experience will be enhanced throughout their entire stay, leading to higher customer satisfaction, reviews and return rates (Metcalfe, 2014; Walsh, 2013).

IV. CHALLENGES AND RISKS

The benefits of using Big Data analytics show the competitive advantage available to SMEs within the travel and hospitality industry. However it is important to be aware of the challenges and risks associated with gathering and handling such large amount of Data. Using the common attributes associated with Big Data analytics, labelled the 6 v’s of Big Data (along with an additional attribute, Visualization, specific to this report), a range of potential risks and challenges will be explored with the intention of making researchers and hoteliers aware of their significance.

A. Volume: The Risk of Having Too Much Data

The travel and hospitality industry may benefit from the availability of large amounts of customer and system generated data, however too much data can be counterproductive and lead to a loss of resources and missed opportunities (Gage, 2014). The risk of analysing such large volumes of data is that the opportunities and benefits gained from Big Data analytics are lost, due to excessive preparation of the data (Ventana Research, 2012). For example, the cost of preparing and analysing the large amounts of data may accumulate to more than the savings its produces.

Strict management and monitoring is therefore required to ensure that the amount of data being analysed is reasonable and producing profitable results.

B. Value: The Risk of Irrelevant Data and Results

Another challenge faced when analysing Big Data is ensuring that both the data gathered and the results its produces are relevant and valuable to the organization. As mentioned previously, this report acknowledges the importance of effective marketing, high quality and personalized services, and effective management of business finances. The data and results it produces must therefore relate specifically to these priorities in order to be valuable to the organization (Tee, 2014).

The challenge faced by analysts however is identifying the reliable and valuable data in a data-rich industry. Consistent and specific priorities and requests must therefore be made across the various services within the hotel to ensure the results and further actions are valuable and effective. For example, marketing employees must work alongside financial managers to ensure that the trends identified reflect on both marketing activities and changes in room prices and budgets (Toedt, 2013). Valuable analysis requires commitment from all services and management levels, alongside a secure and accessible system which allows for the results to be accessed by all, at all times (Greengard, 2013).

C. Velocity: The Challenge of Analysing Data against Time

The challenge of analysing large amounts of data to produce valuable results, also extends to ensuring that the data is analysed swiftly enough to be used effectively (Speigel, 2014).

The challenge analysts must overcome is that an increasing number of bookings are made on the same day, using mobile devices (Bartnick, 2014). This means that the data must be collected and analysed quickly, in order for an advantage to be obtained and used effectively. This also supports the need for an effective and cross-service system to be installed to ensure that access to the data (both inputting the data and visualizing the results) is available at all times, by all employees.

D. Variety: The Challenge of Analysing Various Data Forms

When gathering and storing Big Data, it is also important to acknowledge the two types of data available: structured and unstructured data.

Structured data refers to traditional data stored in databases and typically generated for/by an in-house system. For example, customer booking details such as demographics and credit card details. Unstructured data refers to less traditional data which is typically user/customer-generated and accessed through third-parties and social media sites. For example, photographs uploaded to Instagram or statuses posted on Facebook.

Both forms of data are valuable and important to the analysis process, however each require different skills and tools to analyse therefore increasing the time, effort and costs associated with analysis (Rlijmenam, 2013).

E. Veracity: The Risk of Gathering Unreliable Data

When exploring the different forms and types of data, it is also important to evaluate the reliability and identity of their source. For example, whether the data was sent from a mobile device or accessed through a reputable third-party business.

Research indicates that 3 in 10 organizations believe that their data and following research are accurate; this therefore requires consideration (Ventana Research, 2012). The potential competitor advantage and benefits discussed above rely heavily on accurate and reliable results. For example, forecasting occupancy levels and managing budgets requires accurate initial data to show reliable results. Special attention must therefore be paid when choosing third-party data providers and training must be provided to all involved in inputting and analysing new data (Crandell, 2012).

F. Variability: The Challenge of Analysing Varying Data

When considering the volume, value and velocity of data it is also important to consider the rapidly changing nature of data and the challenges this presents.

Data collected about specific customers is unlikely to change significantly (for example, the demographics of a customer) however collective trends and relationships may change (Bartnick, 2014). For example, a few years ago the concept of using mobile devices over traditional methods when reserving rooms seemed unlikely, however recent research predicts otherwise (Fitzgerald, 2014). Therefore constant management and monitoring of gathering, analysing and interpreting result is required for identifying any changes relevant to the priorities explored in the above.
G. Visualization: The Challenge of Clearly Presenting Results

As mentioned throughout the report, it is important to the data is accessible to all, at all times. However what is more important is that the data is accessed and displayed in a meaningful way. A finance manager may be able to understand correlation statistics, but a marketing employee or kitchen attendant may not. The system installed must provide a way of showing results in an appropriate and meaningful manner, such as simple graphs or side notes attached to statistics (Greengard, 2013). The results must also be interpreted in a meaningful way by all involved, to produce any valuable actions. For example, if a relationship is found between phone bookings and higher restaurant bills, marketing must ensure that they target these customers with relevant promotions. Failure to interpret data correctly could lead to missed opportunities.

V. CONCLUSIONS

Big Data analytics could potentially give SMEs within the travel and hospitality industry a strong competitive advantage, allowing for increased management capabilities and enhanced customer experiences. This advantage exists due to the flexible nature of SMEs, however it could easily be lost if the challenges and risks explained above are not carefully explored and overcome. Hoteliers must ensure that their employees work effectively together to ensure an integrated and effective analysis system is installed, and used effectively, in an industry with such Big Data analytical potential.

VI. REFERENCES


Abstract—With the influx of data over recent years, the use of big data is becoming more common with organizations. However, with there being more data available than storage to contain it, there are issues combating its security and handling. This paper has identified the current problems that exist regarding this matter, more specifically how it is affecting Small and Medium Enterprises. By identifying the issues that arise with this it was possible to propose the appropriate frameworks that are required to be implemented into a company’s infrastructure to combat this.

Index Terms—Big Data, Security, SMEs, Volume, Velocity, Variety

I. INTRODUCTION

Security is an issue that is just as important now than it has ever been before, arguably even more so. In an age where a person’s identity is stored on a system on the other side of the world and accessible online, security is an imperative measure that needs to be stabilized to protect the delicate nature of the data being handled. With ever more data existing on the internet, 1.2 Zettabytes to be exact which equates to 1.3 trillion gigabytes (Andygreenshaw, 2014), it’s becoming more of an issue to handle this amount of data. Threats are constantly apparent with danger of hackers, corruption, and mishandling of information becoming more of an issue.

Corporate businesses of today may well struggle with this issue, but it is the smaller businesses that will be looked at to identify how Big Data is affecting them and the security issues that come with this. Small and Medium Enterprises (SMEs) are just as a big a target for security threats than bigger companies, in fact a total of 76% of small businesses experienced a breach of their security. This is an increase from the previous year where 70% of small businesses’ security was breached and 74% the year before that (Figure 1).

II. WHAT IS BIG DATA?

Big data is a term used to describe the growth of data and the availability of the data. Technology has become a centerpiece in this day and age which includes the internet. However with the growing amount of data it is becoming increasingly difficult to process the growing data sets with traditional management systems. In fact the increase in data in recent years is so vast that 90% of existing data has derived from the last 2 years (IBM, 2014). To understand big data better, it was refined into 3 primary dimensions which include Volume, Velocity, and Variety, although there are many variations of the ‘V’s’ that are used to manage big data (Laney, 2001).

Big data is now an advancing trend that many corporations are adopting, something that the likes of Google, eBay, and Facebook, to name a few, have incorporated into their own companies since the beginning and have therefore not required to adapt their technology and management systems to incorporate it (Lohr, 2012).

A. Importance of Big Data

When big data is used to its potential and analyzed accordingly to benefit ones interests, it can become an important tool that can prolong productivity, create more of an impact on the business, and ultimately extend a company’s existence (Davenport, 2013). Increasing amounts of sectors are employing big data methods and evolving their IT infrastructure. This is highly important because companies now require a method of storing the growing amount of data that they obtain. Every area of a company is starting to amass data, retailers are using social media to identify potential buyers, delivery trucks employ GPS technology for tracking, and healthcare monitor patients constantly to identify potential health risks (Oracle, 2013). All of this is producing data and requires to be stored to be analyzed. If it was not for this, the businesses that incorporate this infrastructure would not be as effective as they currently are.

III. THE CURRENT PROBLEM

The security issues that face big data are not limited to one singular factor. There are a range of reasons why there are such concerns regarding this area be it the lack of knowledge regarding big data itself, or the reluctance to invest in counter measures by businesses themselves. PWC (2014) reports in the Cyber Security Technical Report, that a massive 76% of small businesses experienced a breach of their security. This is an increase from the previous year where 70% of small businesses’ security was breached and 74% the year before that (Figure 1).
These consistent high figures underline the extent of the current problem but is enough being done to address this? In fact the majority of SMEs are only contributing between 2% and 5% of their IT budget on security (Figure 2). This is a decrease from the year before with even more small business not spending anything on security (PWC 2014). From identifying the problem in further detail it will be possible to determine what measures could be taken by SMEs to minimise the threat of cyber-attacks putting sensitive data at risk. However, a more in depth look into why security is such an issue in the first place will help determine what measures are most appropriate to take.

A. Lack of big data knowledge

Businesses are already preparing for the future and are anticipating bigger growth in the data field. There is also a growth in the number of businesses that are incorporating the likes of social media to expand their marketing capabilities. This is a big step for a small business with the amount of incoming data due to inevitably expand, however this has not prompted much in the form of a sound strategic plan to handle this. A report composed by Infogroup (2014) discovered that where businesses are investing more on big data driven market initiatives, 60% of these businesses are not investing in hiring the staff with the knowledge of how to handle the data. McRae (2014) suggests that without having the people with the knowledge of how to make sense of big data, it is meaningless.

In the year of 2013, the majority of incidents that occurred were subject to staff related incidents (PWC, 2014). Without the correct knowledge of data handling, businesses are prone to such incidents. The cost to fix a security related incident, caused by an otherwise avoidable staff error, could ultimately be more expensive than it would have been to invest in educating staff (Olavsrud, 2014).

B. Security threats

A company’s lack of knowledge of the subject is not the only reason for the insecurity of data being handled. The main issue is the direct threat, breaches of privacy and attacks on security systems. Threats can be in the form of anything from a virus on a computer, or a hacker stealing private information. A report by EMC (2014) states that 91% of breaches resulted in data being compromised within mere days, and of those breaches, 79% took weeks to even discover that a breach had taken place.

The report goes on to say that the attacks are becoming more advanced and more intelligent, what with the amount of data out there in a variety of forms. Whereas attackers are evolving, defences have remained static and have not been able to cope with the attacks. This is due to the fact that big data is opening an array of opportunities for businesses, allowing them to offer their services through the latest technologies such as social networks. This is allowing cyber criminals to discover newer opportunities of breaching the security of these newer technologies. Security breaches are more common within larger organisations due to the fact that they are more visible to hackers and are more likely to hold larger amounts of data. However it is evident that SMEs have adopted security measures that are not as adept and will ultimately fail to identify attacks and in many cases prevent such attacks (PWC 2012).

IV. THE 3 VS OF BIG DATA

Big data is split up into three defining properties that differentiate the dimensions of the data, these are called the 3Vs (volume, variety, velocity) however additions to this model have been proposed in recent times (Rouse, 2013). The concerns regarding big data cannot be measured by knowing the volume alone; it is not until all of these properties are considered that it is possible to derive the depth of the
challenges that are being faced. When incorporating this model alongside the security issues of big data it is possible to evaluate how much of an issue it is.

A. Volume

The volume of data is increasing each year and it is expected to keep on increasing more each year with The International Data Corporation estimating that there will be 450 billion online transactions alone by the year 2020 (Andygrenshaw 2014). Data is not only growing in numbers but also in size. In the past it text files were more common and only measured a few kilobytes, whereas now we have graphics, images, music, and video files that range from a few megabytes to hundreds of gigabytes respectively. It is now also common to have devices automatically creating and saving data constantly throughout the day. GPS in vehicles and mobile phones are constantly collecting data and being saved which will allow a business to judge suitable services for their customers (Soubra, 2012). Whatever data it is, it all requires to be stored for it then to be analysed which is now taking up petabytes worth of storage space.

As with the growth of data, there is similarly substantial growth in the amount of security threats. With a shift of Internet Protocol, from IPv4 to IPv6, a paper compiled by Trend Micro (2014) suggests that this could result in huge ramifications in the form of endless cyber-attacks due to the almost infinite number of IP addresses now available. Measures will be required to minimise the load of threats with the window of opportunity opening up for attackers to exploit.

B. Variety

The introduction of newer applications in recent years has presented the opportunity for an array of different formats that all kinds of data can be compiled by. SMEs are now able to analyse what products and services their customers show an interest in via social platforms and from interactions on websites. To some extent it is now possible to predict what a customer wants before they even want it due to the fact that data is constantly being collected from their behaviour from the web (Soubra, 2013). This technology can only perform with big data, and it is only an example of how the variations of data types has exploded in recent years for what is now an everyday tool for businesses.

Information is more accessible now that it has ever been before with the use of phones, tablets, and even TVs. Phones can now be used to purchase goods from as little as a simple gesture at a Point of Sale. There is now a threat all these devices are prone to becoming victim to cyber-attacks. As technology evolves so too will security threats, spam now more common than ever before, whilst viruses, malware and other tools becoming more available to cybercriminals (Trend Micro 2014).

C. Velocity

Data is now processed in real time as opposed to batch process. Where batch would take time to take delivery of a request or result, real time will display results instantaneously. Real time is vital in these modern times; social media would not be as effective without real time updates such as Twitter. The exchange of information is much more fluid and is vital when the data involved is required quickly in order to be effective such as a Tweet of an event (Soubra, 2013). Such is the velocity, that it is extremely difficult to trace the source of a cyber-attack. With the high volume of IP addresses, combined with the variety of possible attacks and the velocity at which these attacks can occur, it is possible for an attack to occur with any trace of the attack ever happening being non-existent (Trend Micro 2014).

V. FRAMEWORK EVALUATION

There are frameworks in place to guide businesses into taking the most appropriate actions with company dealings, how certain areas should be managed as well as timescales that should be met (Braganza. 2000). For small businesses, it would benefit them greatly to adopt appropriate frameworks, Johnson (1993) states that introducing strategic governance frameworks benefits non-profit organisations in an otherwise competitive environment. The frameworks proposed will be suitable for the benefit of small businesses to keeping their data safe from security threats.

A. ISO 27002-2013 IT Security Standards

ISO 27002-2013 is designed to assist businesses that require a means of Information Security Management System (ISMS) designed specifically for the business and for their specific requirements to prevent cyber-attacks due to a current lack of security in place (BSI ISO 27002, 2013). This particular framework derives of 11 clauses to be considered as to how to control the security of current systems. With social media becoming a vital tool to small businesses for marketing and collecting data it is essential that the data they inherit is handled carefully, and not to expose certain data to the public. Section 12 of the BSI ISO 27002 (2013) refers to the control of transferring software, stating that transferring data needs to be defined and documented. The data that a business collects on a customer needs to be stored securely; primarily personal information needs to secure in order to prevent theft of this information which could inevitably compromise the customer. Sections 13 and 14 of the BSI ISO 27002 (2013) refer to control over both messaging and confidentiality, respectively. This will protect the identities of the students and will keep the communications private between the relevant parties. More specifically regarding secure application services on public networks identifies exactly how SMEs should go about ensuring correct authorization procedures are followed, along with keeping integrity, a level of trust, and liability attached with fraudulent transactions (BSI ISO 27002. 2013).

B. ISO 22301-2012 Business Continuity Management System

This standard puts in place an effective Business Continuity Management System (BCMS) to ensure the correct measures are in place in the event that a disruptive incident occurs and what needs to be done in order to manage this. By monitoring the BCMS, the business is able to continually improve the performance and effectiveness of it. This standard is compiled mainly of 10 clauses that are to be followed to get the effective
results for the organisation. Section 7.3 points out the requirement for the persons doing work under the organisation’s control must understand and conform to the policies in place. This is an important step to ensure that staff members are not putting themselves at risk by not following correct procedure (BSI ISO 22301, 2012). Section 7.5 outlines the importance of documenting information, more specifically sub-section 7.5.3 underlines that any documented information needs to be available if and when it is required and must be adequately protected from potential loss, improper use, or loss of integrity. Section 8.2 has outlined the operation aspects of the organisation, sub-section 8.2.3 relays what must be done in order to assess potential risks and the actions that need to be taken in the event of a security breach (BSI ISO 22301, 2012).

These sections are vital to the continuity of an organisation; small businesses can implement this standard in order to grow their business while keeping the relations and data safe and secure.

VI. CONCLUSION

Big data is present, it is here now and more organisations will adopt this approach. With more data becoming available each year it is going to become more of a struggle to harness the volume of it. Big data has huge potential, it is the future, but right now there are too many windows of opportunity for cyber-criminals to exploit personal information which needs to be controlled. Small businesses can benefit greatly from implementing suitable standards in order to keep their data safe.

VII. REFERENCES


Abstract—The purpose of this paper is to identify key aspects of big data in multimedia, how to manage the data and how to store it in most optimum way. This paper will argue how small and medium enterprises (SME) which are part of technology creation can benefit from the proposed technology. Furthermore governance issues such as security, availability and cost will be discussed. It will then pose a strategy that needs to be carried out in order to store and manage the data for SMEs.

Index Terms—Big data; cloud storage; SME; multimedia, sustainability.

I. INTRODUCTION

Data is a collection of facts and statistics collected together that needs to be processed to become an information. Data is a raw material and it is unorganised. Big data is data that has an enormous volume that it becomes difficult to process using traditional database and software techniques. Big data can create lot of constraints in most enterprise as it exceeds their processing capacity (Cheng and Hu et al., 2013).

Gordon Moore, co-founder of Intel, stated in 1965 that the number of transistors on a chip doubles every 2 years. This was the revised to doubling every 18 months. According to Moore’s law computer processing should be able to cope with the growing data. However, the data density has increased to double every 12 months, which shows that Moore’s law does not apply to current growth of data (Wang and Gao et al., 2013).

Although factual data is significantly growing, multimedia is rapidly growing and becoming difficult to manage. There are many factors that are increasing the quantity of the multimedia data. This can be the video quality enhancements or the duration enhancements on YouTube. Multimedia is becoming a key area that needs to be managed and stored using various optimisation techniques to cope up with the storage (Lazim and Rahman et al., 2012).

Storage is also a key concern for organisations as it is becoming difficult to manufacture traditional hard disk with rare earth metal. The alternative and progressive way to more forward is to use solid state drive. Although solid state drives are efficient however, they are still costly per GB when compared with hard disk drive and they have limited write cycles which makes them less future proof (Kobie, 2014).

This paper will attempt to find new ways for small and medium enterprises to provide most efficient solutions for storing and managing data. The solutions can be used to provide a service or to create an individual product which has advancements.

To help SMEs overcome the challenges of big data Vs, detailed discussion will be carried out to analyse how they can overcome challenges in multimedia big data.

To further solve the answers to the questions, this paper will use Zachman framework. It will justify that it is a valid framework to answer the questions that the stakeholders might have.

II. EVALUATION OF BIG DATA

According to Gartner report (2001) big data is three dimensional. It can be defined using three V’s, Volume, Velocity and Variety. Following section will analyse these V’s in more detail along with the challenges they carry. The questions will be asked in V’s which will then be answered in the multimedia big data sections (Kaisler and Armour et al., 2014).

A. Volume

Volume implies to the volume of the data being stored. It is a major challenge in big data as the data is increasing and it is becoming challenging to store it and to analyse it (Katal and Wazid et al., 2013).

How to increase additional storage?

Why should SME manage and store multimedia big data?

B. Velocity

Big data velocity looks at the flow of the data from various sources such as networks, machines and human interaction with the help of smart devices and social media. This flow is always increasing and the real time data can help business make various decisions that may help them achieve competitive advantage e.g. this data can target individual advertisement based on the interest. If organisations overcome velocity by sampling data, they can overcome issues of volume and velocity (Katal and Wazid et al., 2013).

How to increase the performance of data servers?

C. Variety

Big data variety refers to the different types of data whether it is structured or unstructured. This can be media such as photos, song, movies or documents, spreadsheet and presentation. This variety of unstructured data cause many
problems when storing, mining and analysing the data. Therefore it is important to organise the data to add meaning to it (Katal and Wazid et al., 2013).

III. HOW TO MANAGE DIFFERENT TYPES OF MULTIMEDIA DATA?

Although three V’s was a good approach to analyse the consequences of big data, however, as time progressed analysts think it would be a good idea to add few more V’s to analyse the big data.

The additional V’s include Veracity. This is where organisations ensure that the data being stored is correct along with the analyses performed on the data are correct. It is critically important to have correct data when automated analyses are applied, usually where no human is involved, the data must be correct in order to do that. Every one business out of three does not trust their existing data during the decision-making. Therefore when making a big data strategy it is vital that veracity is considered (Chardonnens and Cudre-Mauroux et al., 2013).

Variability is another V of additional big data V’s. Variability particularly important when sentiment based analysis are performed. Variability defines different meaning of same thing depending on the situation. To understand the context of the data, it is important that organisations consider variability (Chardonnens and Cudre-Mauroux et al., 2013).

Visualisation is considered to be the most complex part of the big data V. It involves presenting vast amount of data in to as easy to read and understand manor. This is usually done by graphs which visually displays the data, making it understandable and readable. Therefore it is important for SMEs to make data presentable to perform analyses and understandable (Chardonnens and Cudre-Mauroux et al., 2013).

Value is also an additional V of the big data. This is to determine the value of the data and the added value of the benefits it will have on organisations. Data itself does not have a value, however, analysis performed on it, turning it in to knowledge adds a value to it (Chardonnens and Cudre-Mauroux et al., 2013).

IV. HOW COST EFFECTIVE IS THIS TECHNOLOGY?

A. Multimedia Big Data

Multimedia is an area that is rapidly taking over the huge share of internet and its data is increasingly becoming valuable. Multimedia is particularly valuable as it can help tell a lot about the changing trends at places, people preferences, events, interests etc. All of this can be helpful to many sectors, such as advertisement or directing users to their desire interest (Zorrilla and Martin et al., 2013).

Multimedia big data is a valuable area as it makes up 60% of internet traffic, 70% of mobile phone traffic and 70% of all available unstructured data (Dermot Garland, 2014).

YouTube is one of the most popular media website where users can view and upload videos. Over 100 hours of video are uploaded to YouTube every minute. Duration of the video is increasing, resolutions is getting bigger. The question is, can storage keep up with the demand? (Www.Youtube.com, 2014).

Following are more statistics to provide the context of how big is multimedia. This will also show the need to manage it efficiently.

- 58 million tweets per day on average (Statisticbrain, 2014).
- In 20 minutes 2,716,000 photos are uploaded on Facebook (Kotenko, 2013).
- 154 billion: E-mails sent per day (Fuentes and Patouni, 2014).
- 87%: U.S. adults whose location is known via their mobile phones (Gross and Rock, 2013).
- 2,000%: Expected increase in global data by 2020 (Leber, 2014).
- 111 Megabytes: video and photos stored by Facebook, per user (Leber, 2014).

Although multimedia is an interesting area, however, it is very challenging as images and videos require extremely sophisticated algorithms for content analysis, which is much sophisticated than structured and unstructured data. This shows that there is real need to develop techniques that manages multimedia big data (Dey and Verma et al., 2013).

Figure 1 illustrates the rate of information created versus total available storage. There is a significant difference between information created versus information stored which raises a critical question, where did the information get stored? This is unclear yet although it is certain that there is a desperate need to find efficient solutions that will allow organisations to meet the demand of increasing data.

According to IDC (2014) information that is created is not always stored. For instance a TV show is not always recorded. This could be due to the lack of storage or there is unnecessary data created.

B. Multimedia for SMEs

Most small and medium organisations are making no use of big data analytics. A study by the SAS organisation estimates that less than 0.2 percent of UK SMEs are using big data analytics, one fifth of all businesses admits that they have very
poor understanding of the big data and issues related to it. This leads to a big gap of data analytics between SMEs and large organisations which inevitably leads SMEs in the cold and they are left out (Wilkinson, 2013).

The stakeholders for the proposed project will be small and medium enterprises who are involved in technology creation. It is beneficial for them to manage the data for multiple reasons. They can provide solutions to manage the multimedia data for organisations. Business can also startup their own multimedia content providing service with advanced techniques to manage the data efficiently. The list is endless, in conclusion proposed project will help small and medium enterprises to efficiently manage and store multimedia big data.

To assist SMEs with big data multimedia, big data V’s will be applied. This will help them use multimedia efficiently.

1) Volume

Over the year’s media quality have improved a lot, from 144p videos on YouTube to 4k resolutions. The resolutions are continually improving meaning they require more and more storage. YouTube has also increased the length of the videos uploaded, from few minutes to unlimited upload. Photos also require more storage then before, mobile devices have improved their cameras by adding more pixels on the photos meaning higher quality photo that require more storage. This also applies to song and any other type of media.

To overcome the volume, SMEs can use various techniques. This includes file compression which reduces the storage footprint by compressing each file. To do this SMEs can use various software that are widely available to reduce the storage space. SMEs can apply this compression to backups and to the storage servers (Shapira and Storer, 2003).

Although data compression has been used by many organisations, data de-duplication is the new concept which is not fully adopted by many organisation. This technique eliminates the duplicate copies of repeating data. This technique reduces the number of bytes, unique byte patterns are identified and stored during the analysis. These bytes are then compressed to further improve the storage utilisation. SMEs can use data deduction technique to take advantage of modern technology to compress more data than most organisations (Klimek and Faber et al., 2013).

To further improve the analytics, technology creator SMEs should set media format standards. This will be done after analysing the most efficient standard that consumes the least storage. For instance MP4 is one of the video format which is deployed by many organisations like Apple. This is a very efficient format yet it takes the least storage. This can apply through various kinds of media to provide most efficient way of storing them. It will ensure that the storage hungry format does not get stored (Nagendra and Misra et al., 2013).

To perform analytics on multimedia, files can be stored in various qualities such as 144p, 220p, 360p, 480p, 720p, 1080p etc. this will enable users to access whichever quality they desire based on bandwidth and will enable SMEs to perform the analytics on the lowest resolution which is 144p.

Another option for SMEs would be to use a cloud storage to store all the multimedia. This will reduce their management effort and short term cost. It will also reduce the availability for SME to start projects using cloud server however, as figure 2 shows it will be very costly in a long run and even as a start-up for some SME. The best solution for the long run would be to setup a storage server with all the management techniques mentioned above.

![Fig. 2. Cost of Cloud Storage Per Petabyte](image)

2) Velocity

To increase the performance of the bandwidth and storage SMEs can explicitly set the file formats that will be the most efficient for the performance. This approach will eliminate the low performing formats.

To increase the performance of the storage servers, SMEs can use solid state drives (SSD) which will boost their performance. There are not many organisations using SSDs which will give better performance advantage to SMEs. SSDs are also energy efficient, which means SMEs can keep more data with less power. However, SSDs has limited read and write cycles which many not be very future proof for the SMEs (Yang and Ren, 2011).

The other option to increase the performance of the storage would be to use raid architecture with the hard disk drives (HDD). Raid architecture increase the performance by spreading data on multiple HDDs. Although it will not be as fast as SSDs and SSDs raid architecture can also be configured, it will still be faster than traditional storage structure and it will be suitable future proof option. However, in the long run SSDs will be cost efficient as they consume less energy and over the period they will cost less than HDD (Wan and Wang et al., 2010).

3) Variety

To efficiently manage and store the media, SMEs needs to set apply complex algorithms to identify the media type based on sentiments. This will be particularly useful when performing analytics on the data sets.

4) Availability
This approach will take longer to be implemented in a current SME that does not have the format standards in place as they have to filter current data. Current SMS can also implement this approach from the newly arriving data and ignore the standards of the previous data. This way they can implement this approach from very start.

Data deduction can be applied to the current data and the new data which increase additional storage meaning this technique can be implanted straight away.

V. CONCLUSION

Multimedia data is an area which is rapidly increasing and there are a lot of opportunities for small and medium enterprises to overcome challenges to succeed. Although managing multimedia is difficult and complex, following data management techniques in this paper, SMEs will overcome the big data Vs.

SMEs should store and manage multimedia big data as it will increase the storage space while making them efficient. This give them an advantage over the large organisations.

Additional storage will be increased using an advanced data deduction tools and by explicitly setting format standards that SMEs must use when storing the data. This approach will increase additional storage which is not saved by most organisations.

SMEs will save the cost when managing and storing the data the most efficient way as it will increase the storage, resulting them to take more data on board.

Although SMEs can hire cloud storage vendors to save the start-up cost. However, following the techniques for storing and managing the data they can save the cost in the long run.

VI. REFERENCES


Big Data Security Challenges In SMEs
How The Big Data Challenges Can Be Mitigated

Feyisayo Obisesan
University of Derby
Information Technology
Derby, England
F.Obisesan1@unimail.derby.ac.uk

Abstract—This purpose of this paper is to highlight the challenges involving how small businesses use and manage big data. There are big data challenges that certain small and medium enterprises face due to a lack of qualified staff in the field of big data. As a result small and medium sized enterprises (SMEs) are continually faced with security issues and breaches. The reason for these security issues and privacy concerns are magnified by the velocity, volume and variety of big data. The current security mechanisms which are in place are often too inadequate as they are tailored to secure small scale static data. By stating the proposed benefits of using big data and the ways SMEs can incorporate added security features, big data can benefit organisations. The identification of government frameworks proposes ways for the SMEs to incorporate strategies that will maximize security.

Index Terms—Big Data, Information security failure, Compliance, Business Continuity, risk management.

I. INTRODUCTION

The use of big data within small and medium sized enterprises (SMEs) can be advantageous if used in the right way. SAP (2012) suggests that SMEs can gain a competitive advantage over their competitors by the effective use and management of big data. However, this is not normally the case as SMEs continually face security issues regarding big data and the limitations of how to analyse data. This is due to the lack of qualified personnel who have the expertise to fully understand and make use of the data to benefit the organisation. MacIlnnes (2011) suggests a major problem that hinders SMEs is that they do not have qualified big data employees. A disadvantage of this is that these organisations are not making use of the vast amount of big data at their disposal. MacIlnnes (2011) indicates that organisations have the raw data for example social media or email marketing programmes that can provide much needed knowledge to enable businesses to understand their customer behaviour patterns to identify trends. The technologies to achieve these goals are available to the businesses however the only obstacle is the lack of big data qualified staff within the businesses.

As well as there being information security threats through a lack of qualified personnel, other threats to enterprises are caused by the employees themselves. Cheng-shan, Gu-yue and Li (2007, p.77) emphasises that a large proportion of security issues within organisations were related to inside crime. Companies neglect the most perpetrators of intellectual property theft, their own employees, (Symantec 2011).

Big data according to McGuire, Manyika and Chui (2012) is large amounts of data that can be collected and analysed to distinguish patterns to make better business decisions. As a result this will become the basis for competition and growth for the companies to enhance their productivity. In addition, big data can significantly add value to the world economy by minimising waste but increasing and maximising the quality of products and services. Big data analytics according to SBA (2012) is the process of examining vast amounts of various data. The purpose of big data analytics is to study the large amounts of data in order for the businesses to collect useful information to uncover hidden patterns and unknown correlations, (Kwon, Lee and Shin 2014, p.387).

II. FAILURES IN INFORMATION SECURITY

A. Privacy concerns

Privacy and security issues in small and medium sized enterprises are magnified by the variety, volume and velocity of big data, (CSA 2012). Large scale cloud infrastructures with a wide range of software platforms spread across large networks can increase the likelihood of attacks to organisations systems, (Bollier 2010, p.23). The current security measures that are being used to prevent such threats from happening are not appropriate. This is due to the security measures being fit for purpose for only small scale static data.

As society becomes more dependent on information technology this increases the chances of threats and security breaches. Freed (2014) states that cyber related threats are still considered to be one of the most areas of concern especially to private industries. For industries to gain benefit from big data they should have the right defence mechanisms in place for the increasingly destructive attacks. These attacks are designed to render organisations systems inoperable and vital customer information can be deleted. Freed (2014) mentioned that South Korea suffered a cyber attack against its commercial and media networks. This destroyed thousands of computer workstations and disrupted the online banking systems. These attacks exemplify the alarming trend in mass data deletion and system damaging attacks. Without the use of important data it is impossible for organisations to gather knowledge from the data produced to be able to analyse the data to benefit the organisation.
Yeh and Chang (2007, p.480) states that 90% of enterprises had suffered security breaches. In addition 75% of these enterprises had experienced business problems as a result of security breaches. This is a disadvantage to the enterprises as they heavily rely on IT to add value to the organisation. Small and medium sized enterprises depend on two information values which are the dissemination of information for innovation success and the protection of information to retain competitive advantage, (Mohr 1996, p.247). Information security threats arise from technical system faults or human and administrative errors, (Jung, Han and Lee 2001, p.489). According to Straub and Nance (1990, p.48) people often misuse hardware, program data and computer services which result in security breaches. The answer to the question is that by investing in people the information security threats will significantly be reduced. Furthermore, information systems risk involves the vulnerability of its assets such as hardware and data misuse. Thus, vulnerability can have a negative effect on the information systems assets, (Rainer, Snyder and Carr 1991, p.193). Therefore, these risks occur when the assets are vulnerable to threats. Weber (2010, p.24) suggests that it is imperative for organisations to keep stakeholders information secure and to be resilient to attacks. However, the fulfilment of customer privacy can be difficult to achieve. Furthermore, the adoptions of privacy enhancing technologies were created to prevent such threats and privacy concerns from occurring.

B. Factors Preventing Organisations Adopting Information Security

There are some factors that prevent small and medium sized enterprises from adopting security mechanisms. Von Solms and Von Solms (2004, p.372) suggests that small and medium sized enterprises do not value security measures that are in place. The benefits of security are not considered as important within organisations until a security breach has taken place. This shows that the organisations show a sense of false satisfaction to security measures. Consequently, this prevents those organisations from adopting information security procedures. Failure for organisations to deploy information security is also because of a lack of knowledge given to the managers to perform their responsibilities properly, (Von Solms and Von Solms 2004, p.375). Organisations do not empower the information security managers with the tools; infrastructure and the support mechanisms that they need to enhance security. Instead these employees are left with no knowledge of information security resulting in them either leaving the job undone or doing the job inadequately. As a consequence, the organisation is open to severe risks as the security plan never gets fulfilled.

It is vital that organisations have adequate information security structures to enable a successful information security plan, (Von Solms and Von Solms 2004, p.374). This relates to the way that information security is structured and organised within the enterprise. The benefit of this is to ensure that the regulations of best practices of information security management are met. This suggests that the existence of proper organised security structures include information security mechanisms which are essential for successful information security implementation, (Von Solms and Von Solms 2004, p.375).

III. BIG DATA CHALLENGES

A. Value

Data value is a very important factor for organisations as the data that is being produced must be relevant to the organisation. Loukis, Puzalos and Salagara (2011, p.129) states that data valuable to an organisation can be transformed into useful business analytics.

Demchenko et al. (2013, p.50) defines big data as the added value that the collected data can bring to organisations. The problem that organisations face regarding the value of big data is that these organisations do not have the technology in order to retrieve valuable information that will benefit them, (Demchenko et al. 2013, p.49). With the large amounts of data being produced it is very difficult for organisations to extract meaningful data. Katal, Wazid and Goudar (2013, p.405) identifies challenges that organisations face is filtering the most important data from all data collected to add value to the business. Consequently, systems that can extract important data from large amounts efficiently and effectively will need to be designed.

B. Variety

Chen and Zhang (2014, p.5) states that the vast variety and different kinds of data that is being produced is a challenge. With vast amounts of data it is difficult to efficiently represent, access, and analyse unstructured and semi structured data. Chen and Zhang (2014, p.6) states that variety issues with big data relate to inconsistent and incomplete data being produced and analysed. If the data that is being analysed is untrustworthy then the results will reflect that. Large amounts of data are being produced from many different categories from various sources. For example, from web pages, web log files, emails and social media sites, (Katal, Wazid and Goudar 2013, p.404). These various types of data, structured, semi structured or unstructured are totally different. Consequently, the data becomes increasing hard to manage and to create the valid connections.

C. Variability

Data loads become increasingly challenging to be maintained due to the inconsistency of the data flow, (Katal, Wazid and Goudar 2013, p.404). Due to the varying types of data that is being produced from social media sites, for example, causes peaks in data loads with uncertain and unpredictable events occurring, (Katal, Wazid and Goudar 2013, p.404).

D. Volume

Volume of big data relates to the amount of big data that is being produced, (Buhl et al. 2013, p.65). In addition, because of the large amounts of data that is being produce coupled with privacy concerns cause challenges to small and medium sized enterprises, (Buhl et al. 2013, p.66). Akerkar (2013, p.104) describes big data as large amounts of data that comes from posts on social media sites, digital pictures, intelligent sensors
smart phone GPS signals and purchase transaction records. Bell (2013) refers big data as datasets whose size is beyond the ability of database software tools to capture, manage, store and analyse. Therefore the challenge with the volume of big data is how to deal with size of the data. More than 80% of today’s information is unstructured and is too big to manage effectively, (Akerkar 2013, p.108).

The novel use of the big data V’s highlight peculiar ways in which the big data challenges are portrayed in organisations.

IV. MITIGATING FACTORS

In order to minimise the issues relating to big data in information security a number of frameworks could be used. The use of frameworks identifies the issues and provides a way in which those issues can be mitigated. Information security issues within SMEs happen due to lack of training provided to employees. In addition, some of the employees of the organisations may be the individuals that leak important information about organisation. Cheng-shan, Gu-yue and Li (2007, p.77) indicates that employees within organisations can contribute to leakages of important information. These issues need to be addressed so that the organisations can benefit from the data that they are producing. This data can then be analysed for the businesses to uncover patterns that will enable them to gain advantages of their competitors.

A. Zachman Framework

The Zachman framework provides structured methods that enable enterprises to view and define their enterprise. This framework can be used to organise and analyse data produced by those enterprises. Zachman (2003, p.4) suggests that applying this framework within an enterprise identify this information:

- Identifying someone who has undertaken to do business in a particular industry.
- The business people who run the organisation.
- The system analyst who wants to represent the business in a disciplined form.
- The designer who applies technology to solve the problems within the enterprise.
- The builder of the system.

The enterprises that apply the Zachman framework will be more suited to reap benefits from big data as they will become more adaptable to change. Zachman (2003, p.4) suggests that organisations have great difficulty in responding and adapting to change. This difficulty is due to a lack of internal understanding of the employees about the structure of the organisation. The use of the Zachman framework provides a way that classifies organisations architecture. Furthermore, this tool can reduce information security issues in organisations as it can be used to model current functions, elements and process.

B. Risk Management Framework

The importance of organisations applying this framework is that it can allow the organisations to improve their internal control systems. In recent years there have been concerns on risk assessments and that there is a need for an approach that manages risk, (COSO 2004, P.5). This robust framework effectively identifies, assesses and manages risks. This would benefit organisations that continually face information security issues.

C. Governance strategies and Compliance

ISO (2005) states that maintaining confidentiality; availability and integrity are the primary goals of information security. It is imperative that organisations maintain confidentiality, availability and integrity of data to improve the security of the organisation. ISO (2005) states the importance of data that an organisation produces is like an asset that needs protection. This essential information needs to be protected as it holds valued details that the organisations needs to operate. (ISO 2005). It is also important to protect the information in an increasingly interconnected business environment. As a result of the increase in interconnectivity, important organisation information is exposed to a growing number of threats and vulnerabilities. Moreover, information security is the protection of information from a wide range of threats. This aims to ensure business continuity, minimise business risk and maximise business opportunity. (ISO 2005).

1) Confidentiality, Availability and Integrity

Calder and Watkins (2005, p.73) highlights the importance and the necessity of organisations deploying security measures as it aims to reduce the chances of security breaches. Furthermore, Calder and Watkins (2005, p.74) identifies confidentiality as ensuring that the business information is only accessible to those who are authorised to access it. The protection of the accuracy and completeness of information and processing methods by the protection against unauthorised modification relates to the integrity factors. This would deter unauthorised people from making changes and alternations to the organisations data. Allowing only authorised people to have access to the organisations information conforms to the availability of information governance principle, (Calder and Watkins 2005, p.74-75).

D. Business Continuity

As stated earlier that with the reliance of technology within organisations it is almost inevitable that security breaches will occur, (GAO 1999, P.4). However, business continuity code of practice allows organisations to plan for and respond to business disruptions. As a result, the organisation will continue normal business operations at an acceptable level, (ISO 25999 2006, P.7). Moreover, organisations using the business continuity management process are able to identify potential threats and the impact that those threats can have on the organisation. Having identified the threats, the business will adopt the necessary frameworks for building resilience with the capability for an effective solution that safeguard the interest of the organisations reputation, brand and stakeholders, (ISO 25999 2006, P.8).

Business continuity management is strengthened with adopting the risk assessment management framework as it sets out to understand the risks that organisations face, (ISO 25999 2006, P.12). The risk management framework aims to manage the risks involving services or products that the organisations
provide. The services and products can be disrupted in many different ways which can be difficult to predict. By focusing on the impact of disruption, business continuity management identifies the areas which the organisation depends on for its survival. Through the use of business continuity management the organisation can recognise what is needed to be done before an incident occurs to protect its stakeholders, technology and information, (ISO 25999 2006, P.13).

V. CONCLUSION

Managing information security threats related to organisations’ growing reliance on information technology is an ongoing challenge. Numerous organisations have struggled to find efficient ways to ensure that they fully understand the information security risks that are affecting their operations and implement the appropriate controls to mitigate these risks. These information security risks are preventing organisations benefiting from big data and big data analytics.

After analysing the many different big data challenges the use of frameworks and compliance with governance strategies provides the mechanisms in place to enable businesses to reduce and to fully eradicate information breaches to enhance their organisation.

VI. REFERENCES


Abstract—Big Data have become a torrent flowing into every area of the global economy. It’s the whole new opportunity that every business in any sector should take advantage of. Retail industry is one of the sectors that could really benefit by increasing efficiency and profitability in supply chain execution, with the right use of this data explosion. This paper aims to deliberate what big data is, expand on what benefits it has to offer in the retail supply chain in addition to further discussing the challenges it proposes.

Index Terms—Big Data, Retailers, Supply chain, Benefits

I. INTRODUCTION

Half a century after computers entered mainstream society, the data has begun to accumulate to the point that something new and special is taking place. The sciences like astronomy and genomics, which first experienced the explosion in the 2000s, coined the term “big data” (Cukier and Schonberger, 2013). The concept is now migrating to all areas of human endeavor.

There is no rigorous definition of big data but it’s meaning could be condensed into this: Big Data defines a situation in which data sets have grown to such enormous sizes that conventional information technologies can no longer effectively handle either the size of the data set or the scale and growth of the data set (Ohlhorst, 2012).

Companies all over the world capture trillions of bytes of information about their customers, suppliers and operations by using millions of networked sensors that are constantly being embedded in the physical world in devices such as mobile phones, computers, smart energy meters, automobiles and industrial machines that are able to sense, create and communicate data in the age of the Internet of Things (McKinsey Institute, 2011).

II. WHERE IS THE VALUE?

However, extracting value from all this huge amount of data is definitely not as easy as it may sound. Big data is full of challenges, ranging from the technical to the operational, both of which can derail the ability to mine value and find out what big data has to offer.

Big data has four basic dimensions that are related to the primary aspects of it:

A. Volume

Big data comes in one size: large. The companies are already handling an enormous amount of data containing terabytes or even petabytes of information and according to a recent IDC survey the volume of data that will be under management by 2020 will increase 44 times over 2009 levels (Beulke, 2011).

B. Variety

Big data takes into account two types of data: structured which fit easily into existing databases, such as transactions, conversions or other numeric values and unstructured data which is more amorphous like tweets, the likes on facebook, text, video, audio and log files. The second type of data has to be standardized and ordered before it is ready to be analyzed (Dusto, 2012).

C. Veracity

The massive amounts of data collected for Big Data purposes can lead to statistical errors and misinterpretation of the collected information. Purity of the information is critical for value. (Ohlhorst, 2012).

D. Velocity

Being most of the times sensitive to the time component, big data should be used as it is streaming into the company in order to succeed the maximization of it’s Value to the business. It is also necessary to be available afterwards from the archival resources for future use (Ohlhorst, 2012).

In today’s data explosion in terms of volume, variety, veracity and velocity handling them alone is not enough. The key is the value creation, by analyzing these massive data sets and extraction of data intelligence, in order to have a successful strategy execution.

Nevertheless, the complexity of big data does not end with just these four dimensions. There are other important factors at work as well, like the technologies and processes that are used to define the value of the information hidden in the data sources.

The most important categories that these technologies and concepts are being used in order to create value are the following:

- Traditional business intelligence (BI)
- Data mining
- Statistical applications
- Predictive analysis
- Data modeling

The preceding categories constitute only a taste of where big data is headed or why it has intrinsic value to an enterprise (Ohlhorst, 2012).

III. BENEFITS AND CHALLENGES

A. The Retail Industry

Since the early beginning of time until today, the retail industry has always been a dynamic domain, characterized by hyper - competition. Today, big data offers a whole new provocative dimension through which this competition becomes even more intense as the ways that this new wave of information can benefit the retailers are getting more and more. With this in mind, every retailer who respects his business should adopt and incorporate in his operations big data technologies and concepts in order to remain competitive and successful. McKinsey Global Institute declared that there is a 60% potential increase in retailer’s operating margins with big data (MGI, 2011).

In more specific terms, according to a survey by retail research firm Edgell Knowledge Network, 57% of retailers say that they have or are planning a big data strategy for their business, 30% say that they have already completed a big data project and 13% say that they don’t understand or don’t have the necessary resources to apply it into their business (Dusto, 2012). This means that all the companies, including the SMEs, have recognized the influence of Big Data and have taken it into account knowing that they will either take advantage of it or they will be left behind by their competitors.

In addition, according to retailers that do have a big data experience 46% of them believe that the volume of data with which they must contend is the biggest challenge in dealing with big data, 34% say that managing data variety is the top challenge and 20% say handling the data’s velocity, or the frequency at which it is generated and captured, is hardest (Dusto, 2012). It is obvious that each company, depending from its size and operations, needs to pay more attention to a different factor of Big Data.

B. Supply Chain Management

Big data can be used in retail in many functional areas such as in marketing, merchandising, operations, new business models and supply chain – logistics. This paper focuses on the benefits for the supply chain which is a critical component of a retailer’s business.

The revolution in big data offers a huge opportunity to supply chain managers to start thinking about what data they need to collect and how they will use it. One of the advantages of big data technology is that you can use unstructured data to take notice of things like quotes that never became orders, all the searches that never actually went through to a quote and different buying patterns (Deverajan, 2014).

There are four basic levels of analytics, where big data can contribute, in the supply chain risk management.

- Descriptive - In order to stay in touch with what is happening, many companies use systems or control towers which are able to monitor information from various sources about occurrences around the world such as weather events, news reports, political events and social media tracking. For example, the knowledge of an outbreak of flu can be used to determine which areas may need more chicken soup or cough drops, while details about road construction or disasters due to weather conditions can be used to reroute shipments in order to make sure that the products will reach store shelves on time. Not all from the information that is collected is important for the company, so the control towers have to use analytics to find out the ones that are valuable based on the location of the company’s products, its suppliers, routes. It is crucial to know which occurrence needs to be taken into account and which can be ignored. (Simchi-Levi, 2014)

- Diagnostic. When problems occur in the supply chain operations an analysis needs to be made in order to find out the source of the problem to prevent it from happening again. All the data in the systems are used for this analysis to be completed successfully. (Simchi-Levi, 2014)

- Predictive. There are many risks that could be predicted like changes in demand, delays in the production or shipment. This can be succeed with the help of supply chain analytics tools like network design or inventory optimization which are able to build a supply chain network which will sustain these risks. (Simchi-Levi, 2014)

- Prescriptive. All the collected data about the company’s suppliers, customers, components and facilities could be put in use in order to find out and plan for what happens if something fails even before it does happen. By estimating how it will take to recover from a site failure and the financial impact of this, the company would be able to know where it is vulnerable the most and protect itself. (Simchi-Levi, 2014)

These levels and theory can be put in action in the supply chain management through many ways, which will be discussed from now on.

To begin with, accurate forecasting is at the heart of all business decisions but remains the Achilles heel of the supply chain. The average forecast error for consumer products companies remains 48%, despite the huge amount of information about product movement that is ready to be used (Byrne, 2011).

Large retailers have begun analyzing data from throughout the supply chain using big data analytics to sense demand, by having a real – time view of the product demand, product sales, location of demand and sourcing process. The benefits of this are tangible and financially significant as it could lead to a reduction of forecast error by 40% or more. The most common way companies monetize this improvement is to eliminate production and carrying costs of inventory while maintaining or even improving service quality. Having the product the right moment at the right place avoids expensive transshipments and expedites or unplanned manufacturing changeovers, when
succeeding the customer’s satisfaction at the same moment. (Byrne, 2011).

Improved order picking is another way in which big data can enhance a retailer’s supply chain processes. Order picking is a labor-intensive process. When the orders are picked faster, they can be shipped and therefore be available in stores faster resulting in better order fulfillment. Big data solutions allow data from various sources such as orders, warehouse layouts, historical picking times and product inventory to be analyzed together based on rules defined by the supply chain manager to improve the overall picking process. The results could also enable running the order picking process in simulation mode, which is very useful as by changing various parameters and settings the manager could find out the best practice before rolling out the final improved process to the warehouses and stores (Mehra, 2013).

Another process that can take advantage of big data is the vendor management. Even small-medium retailers cooperate with multiple vendors in their supply chain. These include drop ship vendors, 3PL (third party logistics) vendors, transportation vendors and packaging vendors. Big data analytics solutions enable the manager to compare the vendor’s performance with a set of key performance indicators, which he has defined, in order to ensure that the quality of service and the profitability of the vendor stay at the desired levels. These KPIs include vendor profitability, on-time service and customer feedback and complaints. The KPIs can be tracked in real-time by integrating with vendor systems, financial inputs like costs of products, social network feeds related to vendor deliveries and product packaging. After all, vendors can also benefit as they know exactly what the retailer expects from them in order to continue their cooperation (Mehra, 2013).

Personalized or segmented supply chain is yet another opportunity that big data offers to retailers. Nowadays, the shoppers demand personalized customer service. Big data allow the managers of the supply chain to analyze customer interactions across all channels: social, mobile and web to determine the way the customer is using the products he bought. For instance, a segment of shoppers could get products that are organic, or another segment could be offered value-added features like home delivery. All this can result in increased overall revenue and profitability by simply matching in real-time the available goods with the target customer segments (Mehra, 2013).

Finally, automated product sourcing is a practical way to improve your supply chain by using big data too. Losing revenue due to an out of stock product is painful. Big data tools help overcome this challenge by making possible a real-time view of the product demand, product sales and sourcing process. These tools analyze purchase history, product lead times and many other factors that can influence a product sale, like a marketing campaign, weather changes or strikes. When multiple warehouses or ship vendors are available, the manager should be able to work with data from other locations in order to find the best way for the products to arrive on time (Shaclett, 2013).

The analysis of the collected data, for the improvement of the preceding processes, couldn’t be achieved without the big data technologies and tools. The most crucial of them, which are the real “difference makers”, are the following:

- **RFID tags.** Radio frequency identification chips are a big hype, which has a huge potential to change how supply chains work. They allow to track the objects (products) in real time. This possibility of collecting the real time data on packages, trucks, parts and so on and of analyzing it can lead to big improvements in the supply chain management as all the flaws, lags and problems will be possible to catch “on the go” and thus adjust for them (Semenov, 2013).
- **POS data.** Point of sale data stream can benefit the supply chain too. It can provide a real – time demand signal with price information and price optimization can be fine-tuned with real-time POS data to optimize the profitability (Sehgal, 2012).
- **GPS-based location services.** GPS – based shipment management is entering mainstream, so the retailers have to leverage this technology to dynamically manage supply-side of the supply chain. For instance, by combining the RFID for cold-chain perishable products and real-time GPS location, efficiencies can be gained to reduce the goods damaged due to temperature changes or expiration dates.

**C. Retailers who lead the way**

Many large retailers have already used big data solutions in their supply chain, as they recognized the potential behind them, with Walmart, P&G, Apple, Tesco, Amazon and Merck leading the way. For example, in the early 1990s, Walmart formalized its Retail Link system, which provided sales data by item, store and date, to all of its suppliers. This information led to lower merchandising costs for Walmart, and also saved suppliers time and expense in planning their production and distribution (Waller, 2013 and The center for media justice, 2013). Another example is Dell who use big data solutions to analyze real-time feeds from weather reports, delivery trucks and orders to proactively resolve delivery problems before customers are aware of them (Mehra, 2013). Amazon.com uses collaborative filtering to generate ‘you might also want’ prompts for each product bought or visited (Kalakota, 2011). Finally, FedEx, which is a company that handles nine million shipments a day and all the accompanying data, recently decided to apply that data to physical items by creating next generation, first-of-its-kind information service that combines GPS sensor device and a web-based collaboration platform: Sense Aware (Cronin Movers Group, 2014).

**IV. SECURITY ISSUES**

Security and privacy issues are magnified by velocity, volume and variety of Big Data such as large-scale cloud infrastructures, diversity of data sources and formats, and high volume inter-cloud migration. Proper security entails backing up data and protecting data from corruption, and not only
keeping the bad guys out (CSA, 2012). The most important points that need to be taken into account in order to protect your data are the following:

- Access. The companies can easily protect their data by eliminating the access to them. The key is to control access for a start.
- Availability. The key here is to control where the data are stored and how the data are distributed.
- Performance. Higher levels of encryption, complex security methodologies and additional security layers can all improve security.
- Liability. Liability regards to the sensitivity of the data, the legal requirements connected to the data, privacy issues and intellectual property concerns.
- Data privacy is another big concern. Except for the electronic health records, which are subjected to strict data regulations, particularly in the United States, are less forceful (Ohlhorst, 2012).

V. CONCLUSION

In today’s competitive “lead or leave” market place, big data is considered as an oxymoron that offers both challenge and opportunity. With the right strategies to acquire, manage and analyze big data, the companies have in their possession a powerful competitive advantage that leads them to better decision making, therefore it is crucial for them to get familiar with this big data wave. Welcome to the age of big data. With the right math it is a game changer.

VI. REFERENCES


Life Science: Genomic Sequencing In The Era Of Big Data

Gwyn Dafydd Owen Perkins
School of Computing and Mathematics. University of Derby
Derby, UK
g.perkins2@unimail.derby.ac.uk

Abstract— If the cost of genomic sequencing continues to plummet, then the potential for life saving discoveries will be slowed dramatically. Genome sequencing is a victim of its own success, with opportunities for personal medicines and disease eradication on the horizon we must continue improving our IT infrastructures to meet the ever growing demands of this new and evolving life science.

This paper outlines how genomic sequencing and Information Technology’s new buzzword “Big Data” go hand in hand, with new and innovative software and biological hardware technologies that soon could, transport and store the entire worlds digital life, spoken word and written book securely, efficiently and with vastly improved environmental and sustainable qualities compared with current technologies.

Index Terms—Genes, Genomics, Sequencing, Big Data, DNA, FaspTM.

I. INTRODUCTION TO BIG DATA

The phrase “the new oil” (Grimes S, 2013) is now being used to communicate value and provide emphasis associated with big data and big data analytics. Throughout the globe companies and organisations are harvesting huge quantities of data collected from our smart phones, car journeys, the adverts we watch and even our crime and health statistics (Taylor C, 2014). Big Data is not just being collected on us as human beings, governments and organisations are also collecting enormous quantities of data from environmental disasters and phenomena and, for a number of years, the weather (Freeman A, 2014). Much of this data is initially in the form of unstructured text, video, images and audio data sets. Big Data is a term used when integrating these many different sources into one large data set. When analysed effectively and coherently it can provide valuable insight into anything from Earthquake predictions (Perkins R, 2014) and pollution levels, (Intel Free Press, 2013) to directed marketing campaigns targeting individual human beings habits and their trends (Harris D, 2012).

Big Data is working its way into every avenue of our lives and may soon, if it hasn’t already, change the way we as human beings conduct ourselves within this new data driven society (Grimes S, 2013). Governments are starting to understand the importance of collecting and collating these multiple data sets, providing budget allowances for research into this new technology and its analysis techniques. The United Kingdom’s Government has also continued to invest in big data with the Government funded Research Council investing £47m from 2013 – 2015 in over 100 universities to make new and innovative discoveries and promote collaboration (HM Government, 2013).

II. FUTURE OF BIG DATA AND GENOMIC SEQUENCING

The health profession is no stranger to the benefits associated with analysing big data sets, with the National Health Service implementing the Care.Data initiative (NHS, 2013). By taking GP patient data and combining it with hospital records in a central storage and processing facility the National Health and Social Care Information Centre (HSCIC). Thus laying a foundation platform where computers with exponential computational power and complex analytical tools can disseminate and evaluate the data. This enables more efficient management of illness, monitors the spread of disease and creates models to predict and prevent future outbreaks, which in turn help to drive economic growth (Solon O, 2014).

With advances in virtualisation, cloud and grid computing technologies and the recent breakthroughs in genomic sequencing (Goldman N, et all, 2013) and subsequent analysis techniques, DNA sequencing is now seen as feasible and cost effective (Versel N, 2013).

In 2001 the cost and time associated with sequencing a full human genome was estimated to be $100m over 13 years. Currently these figures stand at under $1000 per genome with machines capable of sequencing tens of thousands of genomes per year (Wetterstrand K, 2014) (Young S, 2014). Organisations are now embarking on huge projects to map, store, share and analyse the sequencing data (1000 Genomes, 2010). One prediction for the NHS is that by 2061 the projected spending on health care will be in the region of 16.6% of GDP in England and Wales, compared with 6.8% in 2015, (Wetterstrand K, 2014) this equates to a rise in total spending per head from £1,745 to £9,914 (Appleby J, 2013).

The field of Genomic Sequencing hopes to curb this spending pattern by studying the entire genetic makeup (genome) of organisms; sequencing, mapping and continued analyses. Conducting this analysis on many genomes, aims to better understand the fundamentals of heritable traits and, primarily, how genes work and communicate and so help to prevent and cure disease (Feldman B, 2013). Another innovative goal of genome sequencing is the ability to build personalised medicine specific to the requirements and needs of the individual and tailored to their genomic blueprint. This
allows for the creation of powerful medicines with no side
affects, directly targeting the mutated gene which is deemed to
be the cause of the disease (O’Driscoll A, et all, 2013).

III. GENOMIC SEQUENCING AND BIG DATA REQUIREMENTS

Advances and breakthroughs in genomic science can only
develop if we continue to sequence many thousands, even
millions of biological genomes and upload them into central
storage facilities where different organisations can interrogate
and experiment with the data using detailed analytical tools.
Deoxyribonucleic acid (DNA) is the chemical compound that
contains the genetic instructions to develop an organism
(National Human Genome Research Institute, 2010). A human
genome has between 20,000 – 25,000 genes, comprising of
approximately 3 billion base pairs. In terms of data
requirements this equates to 100GB of data, or the equivalent
of 102,400 digital photos (Feldman B, 2013). The National
Cancer Institute aims to sequence 1 million genomes, with a
whole genome of a tumour and matching normal tissue
comprising of 1TB of uncompressed data. This would amount
to 1 million Terabyte’s of data (1000 petabyte’s/1 Exabyte) that
requires transfer, storage and analysis. It is now less expensive
to sequence full genomes than it is to store, process and analyse
these huge data sets.

To highlight the full scale of the problem; there is an
estimated ~4.6 x 10^30 bacterial cells on earth, with the majority
being uncharacterised. This represents a vast treasure trove of
largely unseen genetics than can be analysed and potentially,
lead to the discovery of novel genes, complete metabolic
pathways and medically valuable end products (O’Driscoll A,
et all, 2013). These data requirements are now overtaking the
professionally accepted predictions of Moore’s Law (power
and memory of semiconductors will double every 18 months
(Mann C, 2000)) and Kryder’s law (“The density of hard drives
increases by a factor of 1,000 every 10.5 years” (doubling
every 13 months) (PC Magazine, 2012)). The future of genome
science relies on the constant generation of sequencing data,
but soon the limiting factor will be lack of facilities available to
store, process, analyse and maintain these truly Big Data sets
(Sboner A, et all, 2011).

For organisations to effectively and efficiently analyse the
subsequent genomic sequence data sets, it must be transferred
to central storage facilities. Here it can be shared via new
innovative technologies such as cloud computing with multiple
organisations and importantly, where the computational power
and/or technology (Grid Computing) is located, conduct the
deep analytical processes required for the life science to
progress and achieve goals (Kristensen D, 2011). BGI (Beijing
Genomics Institute) can currently move 1TB of data per day
for its customers. They noted, that this worked well for 1
genome, but when transferring data from 50 sequences through
the internet it becomes in-practical and takes around 20 days to
complete (Marx V, 2013). Because of these constraints BGI
and other organisations are resorting to the “sneaker net”:
Overnight shipment’s of data laden hard drives via couriers to
storage facilities, this presents huge data risks with the potential
for data to become damaged, lost or even stolen (O’Driscoll A,
et all, 2013) (Marx V, 2013). Most Wide Area Network’s I/O
bandwidth’s and associated hardware are increasingly
becoming the limiting factor, causing major bottle necks when
transferring Petabyte’s of data efficiently and at the pace
proposed by the enormous throughput of data generated by the
sequencing facilities (O’Driscoll A, et all, 2013).

IV. VAVA VOOM – THE DREADED V’S OF BIG DATA

These two issues of volume and velocity supplement the
current understanding of the challenges associated with Big
Data and the fundamental 3 V’s attributed to the successful
implementation and consequent use of Big Data and Big Data
analytics. It must be noted that there are other articles arguing
for further V’s such as Viscosity, Virility, Validity, Value,
(IBM, 2013). But as Laney 2012 (Attributed as the founder of
the theory of 3-dimensional data challenge (Laney D, 2001)
(Laney D, 2012) discusses all of these additional attributes are
qualities of all data, and not defining qualities of Big Data and
that the essential and defining attributes of Big Data are as first
discussed in 2001: Volume, Velocity and Variety (Laney D,
2013).

V. ASPERA FASPTM

The traditional way to move data over Wide Area Networks
(WAN) and Local Area Networks (LAN) is via the
Transmission Control Protocol (TCP) using transfer
applications such as FTP and HTTP (Aspera, 2012). TCP
provides a reliable and efficient data transfer service when
implemented on low-latency and very low packet loss networks
(Aspera, 2012). Because of TCP’s transfer rate control
mechanism, where the sender requires acknowledgement for
every single data packet (Rouse M, 2006), and if the
acknowledgement does not arrive back in a suitable time frame
due to high network latency or packet loss, the sender will
assume the rate at which the packets are being sent is too fast.
Thus the sender will severely reduce the number and frequency
of data packets being sent, slowly recovering the transfer rate
to a suitable speed (Aspera, 2012). This results in very slow
transfer speeds and hugely under utilised, costly and ultimately
inefficient WAN’s especially with Gigabyte file sizes across
un-reliable and long distance international networks.

A. FASP - How it Works

To overcome these transport bottlenecks Aspera developed
and patented a new transport software protocol called Fasp.
Instead of continual optimisation or acceleration of data
transfer, which albeit do work but are still limited by design to
sub 10Gbps networks (Aspera, 2012). Fasp software utilises
the existing infrastructure, commodity hardware and available
bandwidth to eliminate network-layer bottlenecks. It scales
packet transfer speeds over any Internet Protocol (IP) network
path (Aspera, 2013) to achieve maximum throughput speeds
many hundreds of times faster than FTP & HTTP, and can
guarantee quality and delivery times independent of distance,
lateness or packet loss (Intel Corporation, 2013).

Aspera have succeeded by researching and implementing
two new and innovative approaches to data transport
management: One, they have introduced a new mechanism for identifying and re-sending precisely the individual dropped packets. Two, instead of limiting transfer rates Fasp detects and utilises all un-used bandwidth, Fasp transport management protocol will reduce this rate or number of sessions to accommodate any new flows of TCP traffic, then re-evaluate and if possible re-use by adding another Fasp session, this cycle continues until Fasp’s data transfer has completed (Aspera, 2012).

B. FASP – Case Study Evaluation

Analysis conducted by Intel Corporation in co-operation with Aspera, using their new Fasp software and hardware powered by Intel’s flag ship Xeon Processer E5-2600 based system. Using integrated 10 Gigabyte Ethernet and enhanced Input output (I/O) technologies: Intel’s integrated Data Direct I/O technology (Intel DDIO) and Single Root I/O virtualisation (SR-IOV) where able to achieve a little under 10Gbps with only four sessions running with 5% packet loss and network latency at 500ms, they also concluded that this speed can be replicated across WAN’s and LAN’s. Intel noted that Fasp can easily achieve over 10Gbps and is only limited by available network bandwidth and end to end hardware constraints (Intel Corporation, 2013).

VI. FORGET LINEAR! VOLUMETRIC DATA STORAGE - DNA

For millions of years Mother Nature has been using a novel way of storing large datasets within living organisms, DNA. DNA stores instructions on how to make proteins. These proteins determine our individual physical characteristics and are the reason why we all look different from one another (Robert F, 2013). This presents us with a new, emerging and innovative way to deal with our ever growing demand for data storage. Recent research (Brenna M, 2013) has shown storage density’s of 2.2PB’s of data per gram of DNA. In theory, with current digital data storage requirements expected to exceed 40 Zetabytes (ZB’s) by 2020 (Freudenrich C, 2007), it could be plausible to store the entirety of human knowledge, every written book and every spoken word in a few hundred kilos of DNA (Anthony S, 2012).

The benefits associated with storing data in DNA are right in front of us. We are still recovering and sequencing valid DNA from woolly mammoths that died thousands of years ago despite residing in extreme weather environments. DNA is almost invisible, ensuring it cannot be harmed easily and can be transported with ease. It can survive in extreme conditions (-800C to 800C), it does not require an energy source and is fundamentally sustainable and can be stored for many millions of years with no attenuation of data. Un-like present day silicon based linear storage devices which are non-biodegradable, pollute the environment, un-sustainable and decay over time (Shrivastava S, et all, 2014).

A. Current Research – DNA Storage Breakthroughs

In 2012 researchers at Harvard Medical school and Wyss Institute for Biomedical Engineering at Harvard University announced they had encoded DNA with almost 70 billion copies of George Church’s book: Regenesis: How Synthetic Biology will Reinvent Nature and Ourselves (Leo A, 2012). In data storage terms this amounts to 700 Terabytes (TB), or the equivalent of 14,000 50-Gigabyte (GB) Blu-ray discs stored within a single gram of DNA that can be held on the end of your finger. To add context, storing this amount of data on conventional silicon linear hard disks would take 233 TB disk (compressed) weighing in total 151 kilo’s, not to mention the energy requirements to power the drives (Anthony S, 2012).

The team of researchers using ground breaking theory mimicked the way we encode binary bits as magnetic regions on a hard disc platter, and proceeded to store binary values on the chemical bases which make up DNA (Brenna M, 2013).

In order to aid sequencing, and, to overcome previous problems associated with translating binary directly onto DNA where previous efforts caused repetition errors on retrieval, they engineered a 19-bit address block at the beginning and end of every strand of DNA. This allowed large quantities of DNA to be read out of order and then re-assembled via the address blocks (Anthony S, 2012). Because these 19-bit blocks where located at each end of the DNA strand it allowed for multi-way reading, vastly increasing the process time. This was the first time the whole process completed successfully from synthesising the DNA, encoding binary data, sequencing genome, de-coding binary data and data retrieval, in total for 700 TB it took 14 days to complete (Shrivastava S, et all, 2014).

This research was further enhanced in 2013 by Dr Nick Goldman and his team of scientist’s at the European Bioinformatics Institute (EBI). They announced in a press release (EMBL-EBI, 2013) and through a published research paper (Goldman N, et all, 2013) another ground breaking milestone had been achieved in the field of storing and retrieving data from synthetic DNA. EBI reported that by using a new and improved method of encoding binary onto DNA they had raised the storage density to 2.2 PB’s of data per gram of DNA. They announced that instead of simply translating data into binary and then encoding to DNA chemical bases EBI went one step further and translated binary into ternary (0, 1, 2) then proceeded to encode the DNA chemical bases with these strings and added an error correction scheme. By encoding the data multiple times not only did this increase the storage density, it also helped to prevent errors, and could be read back with 100% accuracy (Brenna M, 2013).

B. Costs Associated with DNA Storage and Linear Devices

Currently this technology is out of reach of mainstream data storage firms due to the time and cost constraints of synthesising and sequencing DNA. At the present time it is estimated the cost of storing 1 MB of data on DNA is $12,500, and $220 per MB to decode. Transversely conventional hard disk storage costs equate to $0.08 per MB as of 2010 (Shrivastava S, et all, 2014). It is inevitable that due to DNA’s colossal density, stability, robustness and energy efficiency in the very near future we will be using it as a long term storage medium especially if the research and breakthroughs continue in DNA sequencing and costs continue to plummet. Not to mention the fact scientists have now managed to store data within DNA of a living organism, albeit for a very limited
amount of time. Storing data within our body could be a very secure and efficient way of taking our digital belongings with us wherever we go (Anthony S, 2012).

VII. CONCLUSION

The scientific breakthroughs awaiting genomic sequencing are clear, and so are the fundamental problems associated with it. These obstacles are not only limited to the life science, but throughout the globe we are now producing more data than we can physically and virtually handle. If we continue with research in genomes at the current rate demanded by the science we must implement new and more efficient data storage and transfer technologies.

This paper has shown that IT managers have a new direction when faced with two of Big Data’s challenges, namely, volume and velocity. With Aspera’s new Fasp transport management software vastly increasing on current FTP and HTTP transfer protocols and if costs for genome sequencing continue to plummet it will have a knock on effect of creating its own self-reliable, sustainable and ever lasting storage medium for the entire world, and its been here all along, DNA.

VIII. REFERENCES


Big Data Within SMEs

Treading Lightly

Simon Ranson
University of Derby
Derby, England
s.ranson@unimail.derby.ac.uk

Abstract— Investing into big data is a business venture, and one routinely taken by growing companies. (infochimps, 2014) There are mistakes that are hit time and again by these enterprises (Iyer, 2014). This paper will provide an overview on Big Data Analytics (BDA) with the goal of making the business decision of whether to invest in BDA or not, clearer, as well as point out avenues for further research.

Index Terms— Big data, SME.

I. INTRODUCTION

Big data is quite famous as well as being infamous in some regards. Target’s pregnancy prediction engine (Duhigg, 2012) provided a case of how powerful proper analysis techniques can be, but they are far from the only ones using big data. Not all data collection and interpretation is so morally questionable, an IBM supported platform in Canada streams a massive amount of data from sensors attached to new born babies, and analyses this data to assist with clinical investigations. (IBM Healthcare, 2014)

However, big data is not a perfect asset. It can be very targetable, and devastating in terms of public relations to lose. A credit card company having to phone a customer can be a terrifying experience for both the customer and the company. (IBM, 2012) With the loss of huge amounts of data not being that uncommon, with the losses, often being within the millions of records (Armerding, 2012), big data is also a big vulnerability.

Analytics can also help in identify when such fraud occurs. (IBM, 2012) However, is the philosophy of fixing a problem using big data despite big data causing part of the problem, solid?

Regardless, at some point in their lives, small and medium sized enterprises (SMEs) may seriously consider looking into big data analytics, and how it affects businesses. Big data brings some serious risks, and some potentially large rewards. This paper will outline what those risks are, how to maximise those rewards, as well as possible alternatives to big data that may work better for some SMEs.

II. DEFINING BIG DATA ANALYSIS

Discussing big data is useless without first properly defining the term, and although quite widely understood, an actual definition is not as easy to defend as one might think.

Just how big is “big”? Several hundred Gigabytes tends to be an accepted minimum, but this still isn’t exactly concrete. The important definition lays, therefore, in what big data analytics does, not in exactly how much space big data takes up. It is possible to execute the same analysis in BDA against smaller data sets, so exactly what big data is doesn’t need to be more explained than: a large store of data that would require an analysis engine to make sense of.

BDA is simply an analytical package assigned to interpreting big data and providing meaningful information on the data to the user. It is used to discern patterns within the data, and expose underlying trends within the data. This also suggests that BDA is centralized around the business case, rather than a technological one. One requirement is that results must be computed and returned with a few seconds, not a few hours or days, and considering the size of the data sets involved, distribution of work load is also a requirement. (McGuire, et al., 2012) (Gopalkrishnan, et al., 2012) (Wall, 2014)

A. The V’s

When it comes to describing big data terminology, named the Vs of big data, exists. Many Vs exist, yet they seem to be arguably condensable into just three. They are: Volume, Variety, and Velocity. (Gopalkrishnan, et al., 2012) (Megler & Maier, 2012) For example, the verbosity of data might be considered to be important, text based analysis is more difficult and time consuming than integer based analysis. But verbosity falls into the Variety category, integer values tend to be much less varied than textual inputs. Another example could be the viscosity of data, how likely it is to change over time. This falls into the Velocity of the data. The three Vs are useful in that they easily describe what gives big data its value.

III. COMMON MOTIVES

Looking at what other companies are doing is a useful tactic when planning business stratagem. Investing into BDA for any company regardless of size will come at a cost. Investigating why and how other companies choose to, or decide not to, invest in BDA will provide useful information.

Big data can provide a competitive advantage when it comes to business planning. Proper analytics can unearth a hidden niece market, or expose what customers are likely to purchase in the near future. (McGuire, et al., 2012) Business insight is almost always the primary goal. However, the business case might not be very clear, and might need extensive testing before meaningful results can be obtained.
Another common reason for BDA investments to be considered by SMEs is because the volumes of data that can be analysed have increased beyond current capability. (Gopalkrishnan, et al., 2012) It’s important to note that here the business goal is clear, to analyse the data with a desired granularity without the loss of accuracy or speed.

Preparing to for global business is also a common motive for investing in BDA. (Wall, 2014) Worldwide ventures tend to involve many intricate details, for example tracking fuel consumption and costs for an international courier service.

To sum up, the real motive for investing into any business venture is to increase the overall value of the business. In this instance, this relies on BDA delivering useful analysis to managers who can make strategic decisions. All these motives have one core idea in common; the business is currently expanding and managers require greater analytical capabilities to remain strategically competitive and efficient.

IV. ANALYTICS POSSIBILITIES

Many companies use analytical engines, some to greater effect than others do. This section will quickly look at and learn from a few noteworthy cases.

A. Netflix

In regards to analytics engines, Netflix once posted a prize, named The Netflix Prize. It was, in short, to reduce the “root mean square error” by roughly eight percent. Netflix offered a prize of a million dollars to a winning team, but in the end decided not to implement the winning algorithm because the increase in accuracy was not being worth the cost. (Amatriain & Basilio, 2012) This shows that uppermost accuracy, although ideally desirable, is expensive, and can be impractical.

B. Microlise

The new-data vs archival data validation capabilities of BDA can be incredibly useful when detecting erroneous and suspicious activities. When a truck starts its engine and begins moving outside normal working hours, without any schedule to suggest otherwise, BDA on telematics data is capable of detecting a theft of a vehicle, and alerting managers. This system aided authorities to recover the vehicle within hours. (Claeys-Jackson, 2014)

V. GAINING VALUE

BDA can only be an asset while the data it analyses is still valuable. If there is no data worth analysing, it will be impossible to create value from the data set, this is simply wasting money and effort. This section discusses common concepts around the use of BDA.

A. User readability

The central point of BDA is to take a large amount of data, and discern simple and decisive conclusions from it. If a user cannot simply read and understand the conclusions drawn from the analysis, the entire process is a waste of time. (Gopalkrishnan, et al., 2012) Therefore, it is prudent to specify what exactly is expected to be visible after analysing. Merely collecting the data is only one part of solving the management problem of making decisions from it. Presenting analysis of seemingly useless data as meaningful and relevant information is the net aim of BDA. (McGuire, et al., 2012)

B. Big Data Analytics As A Product

The nature of BDA can also provide useful information to customers. Coupled with a simple UI, BDA can be provided as a service. An example of a successful implementation of this could be comparison sites, or telematics providers. Each are provided data by their customers, which are processed by the company, then useful analysis is returned.

Search engines also deserve a mention here, generating map/reduce queries for almost instant results on massive reserves of data.

C. Big Data Analytics as an aid to Decision Making

Analytics, in general, provides an aid to decision making. BDA has the advantage of using a large dataset, and coupling this with a set of predictive variables, can indicate ways for a business to adapt. In order to make an informed decision, one must know where one currently is. (Gopalkrishnan, et al., 2012)

D. Saving Money

With a previous investment in a large data structure, BDA can provide efficiency benefits that save the business money. The nature of BDA means that requests for analysis are processed incredibly fast. Managers’ time is expensive to a company, choosing BDA to update old analysis techniques can help pinpoint operation issues with the idea of promoting efficiency. (Gopalkrishnan, et al., 2012) (IBM, 2012)

This idea features well in the telematics industry, as operational efficiency of vehicles is paramount of the success of their customer; BDA provides the accurate, almost-instant feedback that shows managers where improvements are possible.

E. Archival Data Operations

Data is generally quite cheap to store. Archival data can be useful to show developing trends, and company development over a long period. Archival data can help validate new data, and is useful when it comes to performing tests against infrastructure or engine improvements. (Cutt & Lawrence, 2008) The ability to keep and use archival data in a useful manner provides a significant advantage. (Claeys-Jackson, 2014)

F. Innovation

Innovating is much easier to do within a SME than a large company is. This kind of agility is greatly paired with BDA, as the assistance with decision making and forward planning allow for a smaller business to steer itself in a more profitable direction. (Wall, 2014)

VI. LOSING VALUE

As it is important to extract value from potential, it is also important to safeguard that value from being lost or value erosion. This section will quickly outline some of the pitfalls.
A. Losing Track Of Useful Data And Tracking Useless Data

No human is expected to gain knowledge from zetabytes of data without the use of analysis packages. Not using all the data at the system’s disposal pulls into question why it was collected in the first place. (Megler & Maier, 2012)

This supports the idea that data is valueless without a proper means to access it. Thinking of a huge library without any reference guides or organisation patterns, any user wishing to sample a book would spend all of their time trying to find the book they need. Therefore the data may as well not exist in the first place. When applying this to BDA, it exposes two potential weaknesses in developed systems: they provide too much information, or that they don’t provide enough. Finding the perfect level of abstraction should be paramount for this reason. (Iliinsky, 2014)

A potential way to minimise this issues would be to provide an adaptive scale of granularity within an analysis package. The ability to compare huge sets of data must be just as possible as the ability to compare several tiny sets, or possibly even individuals. (Megler & Maier, 2012)

B. Note on Vulnerability

Possibly the worst nightmare of any business employing the use of BDA is to have the data lost or stolen. With the dawn of the information age, criminals don’t need to move from their homes in order to cause massive damage, or conduct a heist. (IBM, 2012) Big data provides a metaphorical treasure chest that attracts crackers.

Vulnerable data is not a value to a company it is a liability. It can be incredibly costly to secure an insecure system against relatively cheap attacking threats. The NHS provides an example of how not to do data security, by losing millions of patient records per year. (The Telegraph, 2012)

C. Licencing

Licencing the infrastructure used to store and analyse the data, and licencing any data the business will not collect itself, can be quite expensive. (Gopalkrishnan, et al., 2012) Luckily, there are simple mitigations for these problems: use freely licenced operating systems on servers such as UNIX and free development platforms such a Java.

Another solution could be to outsource these requirements to a cloud service. This reduces costs, because the company will not have to manage or maintain the data structure, but sacrifices a degree of adaptability. Depending on what implementation of cloud used (typically A.a.a.S. , S.a.a.S. , I.a.a.S. , or M.a.a.S. ) would also determine what skillset employees would need in order to develop for that platform, if required.

D. Scaling

Another consideration is the estimation of how data will grow over time, as well as demand for access of the system. Employing a cloud service is incredibly efficient and cheap when comparing to traditional methods of data storage; single servers simply do not have the level of abstraction to be useful when dealing with large calculations that have speed requirements. Clouds also have the benefit of being incredibly agile and scalable.

VII. SMART DATA ALTERNATIVE

Alternatives to big data exist; they tend to require more management than big data. If the business goal is to advertise with complete effectiveness, with absolutely no wasted adverts, the customer must indicate exactly what they want to see, when they want to see it, and how they want ads delivered to them, for any chance of such a system to work. Out of this comes the idea of smart data. (AaronSon, 2013)

Linking back to the earlier Netflix example, absolute accuracy is actually expensive to obtain. Asking each individual potential customer exactly what they want adds a certain strain to registration mechanisms.

A difficulty of smart data is verifying its reliability. Smart data absolutely depends on the trust in the data. The goal of having absolute accuracy is impossible if the supplied request is not properly described, or if the original user made a mistake. Making sense of data is difficult enough, without starting to question the validity of the data itself. (King, 2002)

Naturally, this problem also occurs within big data and BDA, but the end goal of BDA is not to have absolute accuracy, it is to get a clear view of the big picture.

The main difference between smart data and big data is that smart data carries its own semantics, whereas big data has no inherent semantics, it is just there. (King, 2002) The trade-off between big and smart data is therefore bulk simplicity vs accuracy.

VIII. CONCLUSION

Big data is rather a loose term, and its meaning actually does not matter that much. Big data analytics is simple to define relative to big data, but the focus is what BDA can accomplish, not exactly what it is.

Commonly, big data provides an insight advantage. Analytics provide the understanding of trends within big data and should be engineered to be simple to understand and manipulate. It is also possible for BDA to serve as a simple upgrade to existing analytical systems. Preparing for future expansion is also a common reason for considering investing into big data.

It is important to maintain usability of analytical packages, not having access to conclusions drawn from the data makes the entire venture worthless. A simple UI is also advisable when planning to expose BDA to the public or to a customer. The added insight BDA provides assists an agile business keep moving towards greater profits, this is much easier to achieve with an SME than with a larger corporation. Tactical insight also provides an opportunity to boost efficiency.

The value of data collected must be guarded and maintained which can be costly. There are some significant PR and legal risks when it comes to storing large amounts of data, but when properly managed those risks are mitigated. Ensuring that useful data is not lost in a sea of useless data is a requirement of a good implementation.
One of the largest costs will be in licencing operating systems for server machines; it is highly recommended that individual machines run a flavour of Linux to reduce these costs, unless the business plan requires outsourcing. Clouds provide great scalability and agility when it comes to handling large-scale data and applications; their use is highly recommended.

Finally, proper planning and forethought makes BDA an incredibly powerful tool. Getting an implementation that fulfils business needs is paramount, and maintaining the value of your big data implementation comes in at a very close second.

IX. REFERENCES


Moral & Ethical Ramifications Of Collecting Big Data From Consumers

Joe Rawlings
School of Computing
University of Derby
Derby, United Kingdom
J.Rawlings1@unimail.derby.ac.uk

Abstract— Storing customer data for business purposes is not a new technique. However, more and more businesses are employing Big Data sets containing more personal information about consumers so they can deliver a more ‘personalized’ experience, such as targeted advertising, etc. Some businesses, however, have fallen foul of various ethical considerations and have been justifiably criticized. Whilst this might be morally wrong, it may not have been technically illegal, and as such measures have been / are being proposed to ensure businesses collect, store and use data in a transparent, ethical fashion. Whilst consumers are often vocal in their displeasure of businesses acting in unscrupulous ways, it is often found that the same consumers have neglected to read all the privacy and security notices placed on a site before committing to using it. It has also been shown that some consumers are willing to share their information for something as simple as a monetary reward. Consumers and businesses have differing opinions on their definition of ‘privacy’, and as such there have been suggestions as to how to bridge the gap between consumer and businesses. Reaching an agreement on this and having it enforced widely would benefit both consumer and business, as it is now easier than ever for a consumer to voice their concerns (on, say, a social media site) and have them noticed the world over.

Index Terms— Big Data, Privacy, Ethics, Morals.

I. INTRODUCTION

The storing of customer data by businesses is not a new thing. However, with the ever-rising number of consumers connecting to services such as social media sites, this stored data is now being analyzed in such a way to deliver tailored experiences to individual users. Whilst this seems innocent enough, consumers are beginning to realize that their definition of ‘privacy’ differs to that of a business. This paper aims to describe the difference between moral/ethical issues & legal issues, ethical issues from a business standpoint when storing data about consumers, the issue of privacy when dealing with a business that stores data, and the opinions of both businesses and consumers on these issues.

II. ETHICS

It is generally assumed that if a business is going to be handling any personal information about a consumer, they are going to do it in an ethically sound way. This is not always the case, as from a business standpoint something may well be unethical yet perfectly legal. This is not to imply that businesses will forgo ethics purely for the sake of profit, however. Myriad factors, such as increased public awareness of their online fingerprint, have made ethically sound practices more of an issue than they would have been during the dawn of the Internet, and there is the possibility that well-established businesses simply haven’t become privy to the shift in ‘ethical awareness’.

What is considered to be unethical in the Big Data world? Standard ethical procedure applies, which includes things such as:

- Storing and / or sharing personal information without consent
- Making it difficult to, or flat-out refusing to allow the consumer to opt-out of having unnecessary personal data stored
- Penalising consumers against clauses dictated in intentionally long, confusing Terms-of-service documents
- General dishonesty between business and consumer

Specifically with the storage of Big Data however, this question does not have a definitive answer as the general consensus on ‘how far is too far’ is likely to alter in response to public outcry. Big Data is not alone in this field - other ‘socio-technical’ phenomena are susceptible to being likened to ‘a troubling manifestation of Big Brother’. (Boyd & Crawford, 2012)

There have been efforts to attempt to tie together the legal / ethical divide (it is worth noting that the majority of these are on the side of the consumer) and allow for a simple declaration from a consumer as to whether they consent to their ‘personal’ data being stored (name, address and so forth). One such effort was manifested as the U.S ‘Do Not Track Me Online Act’ of 2011. This act set out to do exactly what was mentioned above - allow consumers to explicitly opt-out of having their data stored. The act allowed the Federal Trade Commission to set a strict set of rules for businesses to follow in relation to the above, effectively making it a crime to disobey them. The act also allows for exceptions, such as in a case where a consumer does not want their data storing, yet it is necessary to provide them a service - e.g. parcel delivery will never succeed without an address. In these cases, data is required to be ‘anonymized or deleted upon the provision of such service’. (United States Government, 2011)
Consumer reactions coupled with the desire not to risk alienating said consumer base have lead businesses to rethink the way they currently store sensitive Big Data, and considering the fact that ‘Do Not Track’ was passed, they will soon legally have to.

III. ETHICS & STORING BIG DATA

If / when a SME decides to store data from its consumers, a series of moral and ethical implications come to light. As storing Big Data is now commonplace, consideration to any backlash or furore from users should take place before any data are stored. Perhaps the most famous example of this is when the American shopping chain Target introduced a ‘predictive analysis’ service that was able to identify when a customer was pregnant purely based on a shift in their shopping habits. From a business standpoint this proved exceedingly effective, however the ethical implications here are apparent as the customer is effectively being exploited whilst they are in a vulnerable position – it is suggested that a customers well-established shopping habits may alter during pregnancy due to the extra stress one would experience. (Duhigg, 2012)

An ethical issue to be considered when storing big data is that of persistency. How long will the stored data be kept for? This question is often answered with an ambiguous response at best, as a business’s strategy may change based on the data they are receiving, or a shift in trend in the received data. For example, if you register for a Facebook account and then delete said account, certain data will be retained. This stems even further as if you share information with an app on the Facebook platform, it may also store data indefinitely on a meta-level. (Johnston, 2012)

As the user is given no choice in this matter, this is an ethical issue. However, one could argue that this is how the service can continue to be provided for free of charge. This information is also provided in the terms & conditions upon signing up, which the user should read, but often doesn’t. This may prevent any legal retort, but is ethically questionable nonetheless.

Similar to Target’s predictive service, other companies have attempted similar things with Big Data. One such example is the company OnStar. OnStar provide a service in where your vehicle is tracked via GPS by their systems, allowing for support such as remote diagnostics and stolen vehicle location. In 2011, the company updated its privacy policy stating that if a customer cancels their service and doesn’t explicitly say so, the company will continue to track and store the customers’ data. (Kravets, 2011)

This practice is ethically questionable, and directly pertains to persistency. The company claim that by doing this they were facilitating an easier re-subscription should the customer decide to, however as this data could be stored indefinitely this could negatively impact the customer if, for example, they signed up for car insurance and had their premium inflated due to their GPS history. (Davis, 2012)

It is worth noting that soon after this, OnStar reverted their privacy policy, suffering a negative PR blow in the process. This may seem trivial, but it could easily have irreparably damaged the reputation of the business. (Woodyard, 2011)

This serves to show that consumers are becoming much more aware of what data is going to be stored about them, and as such any ethical implications should be strongly considered before putting systems like these into practice.

All of the above examples of ethical issues within the storage of Big Data do not break any laws. This is an issue unto itself, as whilst something may be unethical or immoral, if it is not illegal then technically the business has done no wrong. Based on this, there has been an argument for creating a series of guidelines a company should adhere to when storing Big Data. This argument states that when storing Big Data, a set of four values should be adhered to – privacy, confidentiality, transparency and identity. (King & Richards, 2014) If these values are respected, then Big Data can be stored in a way that will not encroach on a user’s privacy, and also provide a wealth of useful data for the business.

IV. ONLINE PRIVACY

As more and more companies are storing data about individuals, the actual definition of privacy has altered. Some would argue that in an age where so much data is exchanged online, privacy is no longer possible online. For example, Mark Zuckerberg, creator of Facebook, believes that privacy is no longer a ‘social norm’. (Johnson, 2010)

Likewise, the CEO of Google, Eric Schmidt, believes that only those who have something to hide should be worried about online privacy. (Huffington Post, 2010)

These remarks were met with criticism, as to suggest an individual should allow sensitive information of theirs to be openly monitored if they are not doing anything wrong is controversial at best. The issue further expands if different companies store different sets of personal information on a user, as it means simple mapping methods could be applied to each site to amalgamate a set of sensitive information. An example of this being put into practice is when AOL released anonymous users’ search histories. The user’s name was not given, however with a minimal amount of effort it was simple to correlate the data trail and reveal the name of the user. (Barbaro, 2006)

Whilst the aforementioned cases raise both questions and controversy, they are merely examples of when Big Data storage went wrong. Big Data is stored and utilized to great effect by a vast array of businesses on a daily basis without incident. The following section aims to identify the average consumer standpoint on personal data storage and the standard business case for storing personal data.

V. CONSUMER OPINION VS. BUSINESS OPINION

There has been substantial research into consumer opinion regarding their online privacy. One such piece of research focussed on Facebook specifically and its privacy issues. For a while, Facebook was rated so low on a privacy / security scale it was charged with severe privacy flaws by a watchdog organization, Privacy International. This was followed by many groups being set up - one of which, ‘Students Against
Facebook News Feed’ had over 700,000 members lined up to boycott the company. This clearly indicates that consumers have grown increasingly concerned over their privacy and security on the internet.

However, the same piece of research indicates test cases on Facebook wherein companies have created phoney accounts purely for data mining purposes. One such account was created, and ‘friend invites’ were sent from it to 250,000 people. 30% of these people accepted said invite, showing that whilst consumers may be aware of privacy issues, they do not always abide by best practices. (Bernhard Debatin et al., 2009)

Another piece of research aimed to identify if consumers would be happy to share ‘secondary’ information about themselves for marketing purposes if there was an attached monetary award. Whilst this is not an issue with Big Data per se, it does highlight the fact that some consumers are willing to allay their concerns over privacy for a simple cash pay-out. In some cases, the pay-out was as low as $39. (Hann et al., 2002)

This is a cause for concern, as many consumers who would happily boycott a company for mishandling data would willingly sell theirs on for a price.

This is not to suggest the consumer is always in the right in regards to having their personal information inadvertently stored at a business location. As mentioned earlier, consumers will often forgo reading privacy notices or Terms-of-service documents due to any number of factors. For example, a consumer would be less inclined to read a privacy notice if it was 40 pages long and written entirely in marketing jargon. Whilst it could be argued that this is unethical, the notice is there to be read and as such the law would be on the business’s side in the case of a consumer neglecting to read it.

There has been research into how the perceived comprehension of an online privacy notice affects a consumer’s ability to trust said notice. One such study showed that certain factors dictate how likely a consumer is to read an online privacy notice, and these are as follows:

- Readability
- Noticeability
- Framing effects / layout
- Explicitness of conclusions

The research found that ‘Perceived comprehension of a privacy notice is positively associated with a tendency to read online privacy notices’. Therefore, businesses should make their notices easy to see & read, and draw unambiguous conclusions to avoid confusing consumers. (Milne & Culnan, 2004)

Negativity aside, businesses tend to try and use Big Data for purely analytical purposes rather than malicious mining of consumer information. One such example of Big Data being put to analytical use is ‘now-casting’. Now-casting refers to the use of real-time data to describe activities before official sources are available. This was put into practice by Google to track ‘Flu Trends’, and when the results were correlated with official health reports it was shown that Google’s estimates of flu outbreaks were 97 to 98 percent accurate - not to mention recorded two weeks before the official health reports. This goes to show that Big Data can be a valuable tool and of great benefit to the public. (Bollier, 2010)

Clearly, the consumers and businesses have differing opinions on what constitutes an acceptable level of privacy. It is argued that there are three different perspectives when it comes to consumer privacy: the ‘corporate’ perspective, the ‘activist’ perspective and the ‘centrist’ perspective. The corporate perspective argues that corporations ‘are the primary creator and provider of economic growth and development for society’, and that ‘any restrictions placed on the corporation’s ability to access personal information about consumers only compromises the corporation’s ability…to fulfil its social responsibility’. The activist perspective argues the opposite, in the sense that ‘information will be available to anyone for any purposes…violat[ing] the right to privacy as well as imposing harmful social costs on society.’ The centrist perspective occupies the middle ground, suggesting that the ‘best’ parts of the previous two perspectives can be combined into one, allowing consumers access to a ‘vibrant economy’, with choices made more relevant through the use of ‘reasonable corporate access to personal data’.

As you can see, there are vastly differing opinions on the matter. In an ideal world, an effort would be made for every consumer and business to gravitate towards the centrist perspective, however in practice this may not be possible based on factors such as individual culture, etc. (Bies & Culnan, 2003)

VI. CONCLUSION

In conclusion, the collection of Big Data related to individual consumers is both a help and a hindrance. One the one hand, the quality of data stored is much higher when it can be targeted at specific users / groups of users, as it allows a pseudo-bespoke response to each individual - ranging from targeted advertising through to notifications of health scares in the users’ area. Conversely, the way in which some of these users information is collected has raised controversy, and with more social platforms available to users than ever, expressing concern over these issues is both easy and effective. Many companies marketing wings will now employ Social Media managers, who will log and deal with complaints received over said platforms.

As such, collecting sensitive information from consumers is a gamble at best. Whilst an idea may seem fool-proof from a business standpoint, all it could take is a few users to voice their concerns on a social media website and watch said concerns snowball into something much bigger, or ‘go viral’, and be picked up by various news outlets, damaging business’s reputations.

Once again, this is not an issue with Big Data itself - more the context in which the data are collected. As with any procedure that straddles the line between a business issue and a social issue, there are going to be detractors from both parties. However, with no end in sight for the widespread use of, and reliance upon social media sites, for the time being the
collective voice of the consumer carries much more weight than that of the business.

VII. REFERENCES


Abstract – This paper aims to identify opportunities, challenges, benefits and operations that SMEs face regarding the development and implementation of a big data analytics strategy as part of their organizational structure. The need for the study seemed even more evident after it emerged that only a very small fraction of SMEs admitted to have a big data analytics strategy implemented or, at least, projected. The study starts by evoking the Moneyball episode as a font of inspiration for SMEs. After a number of brief definitions related to big data, the report focuses on the analyses of alternative routes available for SMEs to enlarge and consolidate their analytics’ skillset base as solid path towards building a strong analytics approach. The preliminary results of the research show that, regardless of the adopted approach(es), SMEs will, no doubt, face significant obstacles in their quest for assets that will allow them to achieve that. As a conclusion, the report urges SMEs to carefully evaluate their options before deciding which route(s) to follow while recommends employers and academia to work together in order to further include analytics as part of the latter’s curriculums.

Index Terms – Moneyball, Sabermetrics, Big Data Analytics, Small And Medium-Size Enterprises, Business Strategy.

I. THE MONEYBALL EXAMPLE

In 2002, the Oakland Athletics - the A’s - had one of the lowest payrolls in the most important baseball league of North America - the Major League Baseball (MLB) (Wolfe, Wright & Smart, 2006) having spent less than a third in payroll - US$40 million - than the highest spending team - the New York Yankees (US$125.9 million) (Gerrard, 2007).

After being knocked-out by the Yankees in the opening round of the American League playoffs - the American League Division Series - the season before and having lost some of its best players to free agency, the A’s’ General Manager (GM) Billy Beane was struggling to assemble a competitive team for the 2002 season due to budget limitations (Thomte et al, 2013).

By re-evaluating the club’s strategies, Beane decides to take a different approach in how to bring undervalued but affordable players to the club; instead of the traditional prospection methods where experienced scouts somehow had the final word over which players should be brought in (Cullen, Myer & Latessa, 2009), Beane, facing ferocious internal opposition but with the backup of Economics’ Harvard graduate Paul DePodesta - the club’s Assistant to GM (Lewis 2004, p.17) - decides to follow a more precise analytical methodology - the sabermetric approach (Cullen, Myer & Latessa, 2009).

A. Sabermetrics Definition

Sabermetrics is the term used to describe the analysis of baseball through objective evidence based on baseball statistics (Glass & Lowry, 2008) and was coined by influential statistician Bill James based on the acronym for the Society for American Baseball Research (SABR) (Beneventano, Berger & Weinberg, 2012; Glass & Lowry, 2008; Sullivan, 2004) which have found numerous statistical indicators of player performance and value to a team that pass completely unnoticed to the baseball community (Felder, 2008). In addition to assessing talent, sabermetrics also has the power to influence tactics used during games (Wolfe et al, 2007).

B. Measuring Success

In sport, success is typically expressed in absolute terms such as the total number of medals that a country wins during an Olympic event (De Bosscher et al, 2008). Therefore, in terms of sporting success, Beane’s initiative wasn’t fully rewarded as the A’s failed to reach the summit and win the World Series, i.e., to be MLB champions (Reider, 2014).

VIII. INTRODUCTION

“Small businesses don't realise the tsunami is coming”
- Prof. Philip Treleaven

The statement made by Prof. Treleaven (2013 cited in Cooper 2013) of University College London (UCL) during a panel discussion surrounding big data analytics couldn’t be more explicit in suggesting that SMEs are far from being ready to deal with the emerging technology (LeHong & Fenn, 2012). During the same event he warned small businesses to “protect themselves against a coming ‘tsunami’ of data by investing in big data analytics infrastructure” and blamed a “lack of big data analytics skills for the gulf in supply and demand for key skills in the field of big data analysis” (Cooper, 2013).

Prof. Treleaven’s statement reinforces the need for the study as the aspect related to the topic “SMEs do not know that they do not know” identified and covered by this report is exactly the one highlighted by the lecturer: the existing lack of qualified personnel in the field of big data analytics.

The paper also aims to raise awareness amongst SMEs regarding the importance of having a big data analytics strategy implemented as part of their organizational structure in order to be able to thrive in the current economic climate.
But Beane’s achievements cannot be purely measured by the A’s’ sportive success (or lack of it). Beane, of course, wanted to win but his ambition was far beyond winning baseball games or titles; most of all he wanted to change the game of baseball, to even the playing field (Penzner, 2011).

Beane did, in fact, change the way the baseball community - including teams, managers and executives - would approach the game thereafter and by doing so he became one of the most successful managers in baseball history (Sarkett, 2004).

C. Formula Emulation

The A’s’ success formula sparked widespread interest nationwide and its front office got increasingly busy dealing with enquiries from a cross-section of U.S. businesses and sports entities including teams from National Football and Hockey Leagues (NFL and NHL), National Basketball Association (NBA), Wall Street firms, Fortune 500 companies and Hollywood studios (Wolfe, Wright & Smart, 2006).

Soon after the 2002 season - coined thereafter as the Moneyball season (DuPaul, 2012) - other MLB teams became aware of Beane’s practices and methods and started adopting them for their own (Vito & Vito, 2013). Inspired by the A’s’ achievements, John W. Henry - owner of MLB’s team Boston Red Sox at the time and now the owner of Liverpool FC - failed in his attempt to hire Beane and DePodesta themselves as Boston’s GMs (Pollard, 2012); but Henry followed Beane’s advice and hired similarly inclined Theo Epstein (Hakes & Sauer, 2006), a 28-year-old Yale graduate (Pollard, 2012), making him the youngest GM in baseball history (Shaughnessy 2003 cited in Hakes & Sauer 2006). In addition to this, the Red Sox hired the dean of sabermetrics, Bill James, in an advisory capacity. Henry’s determination and belief in sabermetrics soon paid off: in October 2004 - only two years into Epstein’s tenure as Boston’s GM (Roberto 2005, pp.9-10) - the team broke the “curse of the Babe” and won their first World Series since trading Babe Ruth to the Yankees in 1919. Three years later, another championship would arrive at Boston’s Fenway Park (Cullen, Myer & Latessa, 2009), home to the Red Sox (Roberto 2005, p.9).

D. The Aftermath of Sabermetrics

Nowadays more than half of NBA teams employ tools of analytics, most MLB teams consider analytics a normal part of baseball operations (Alamar & Mehrrota, 2011) and many other soccer, hockey and football teams use performance models with tangible and encouraging results: for example, a study by Biderman (2010 cited in Armstrong 2012) shows that in the first part of the NBA’s 2009-10 season, the 15 teams with at least one full-time statistician on their staff won 59 per cent of their games, while the 15 teams with no statisticians won only 41 per cent (Armstrong, 2012).

II. DEFINING BIG DATA AND BIG DATA ANALYTICS

It seemed relevant at this stage to summarise what is broadly understood by big data and big data analytics.

A. Big Data/Big Data Analytics Definitions

Big data can be defined as data that is complex in terms of volume, variety, velocity and/or its relation to other data, which makes it hard to handle using traditional database management tools whereas big data analytics refers to analysis techniques operated on data sets classified as “big data” (Mouthaan, 2012). However, defining big data isn’t truly consensual with some authors considering that it doesn’t have a clear definition (Dumbill, 2013; e-skills UK (II), 2013; Mayer-Schönberger & Cukier 2013, p.6; Zaslavsky, Perera & Georgakopoulos, 2013).

B. Big Data Characteristics

More consensual is the existence of the three main properties required to define big data also known as 3V’s: volume, variety, and velocity (Zaslavsky, Perera & Georgakopoulos, 2013). These attributes can be described as follows:

- **Volume**: Relates to the size of the data such as terabytes (TB), petabytes (PB), zettabytes (ZB), etc. (Zaslavsky, Perera & Georgakopoulos, 2013);
- **Variety**: Relates to the types of data. Big data extends beyond structured data to include unstructured data of all varieties - text, audio, video, click streams, log files, etc. (Ohlhorst 2013, p.3);
- **Velocity**: Relates to how frequently the data is generated (for example, every millisecond, second, minute, hour, etc.) (Zaslavsky, Perera & Georgakopoulos, 2013).

Some researchers go even further to consider the existence of an extra four attributes - known as the 7V’s - needed to achieve a more broader description of big data (Livingstone, 2013; van Rijmenam, 2013) being Value - although most pieces of data individually may seem valueless, somewhere within that data, there is some valuable information ready to be extracted (Zaslavsky, Perera & Georgakopoulos, 2013) - and Veracity - purity of the information is critical for that data to be considered valuable (Ohlhorst 2013, p.3) - the most significant.

III. MONEYBALL LESSONS

Beane’s determination to apply the sabermetric approach was a remarkable success. Despite his fervent belief in the new system, putting it into place - and therefore be able to take full advantage of it - was probably Beane’s most challenging task as he lacked practical knowledge and expertise. Restless and resolute, Beane hired a team of young people with mathematical and statistical expertise - including Harvard graduates DePodesta and David Forst - to help him fully implement the new approach (Roberto 2005, pp.5-12).

SMEs don’t share the same number of resources or have the same financial capacity of large enterprises (Lu & Beamish, 2001) - likewise the A’s didn’t, back in 2002 - but they can get an edge over the competition if they are able to take full advantage - like Beane did - of an emerging technology (LeHong & Fenn, 2012) that, according to a study by e-skills UK (I) (2013), even the majority of UK’s business giants are still to accomplish. In fact, only 21 per cent of them admitted to have a big data analytics strategy implemented or, at least,
projected. However, and according to the same study, the scenario is even glummer for their smaller counterparts as less than 0.2 per cent of UK’s SMEs confessed to be in that situation. According to Villars, Olafson & Eastwood (2011), being unable to appreciate how analytics can improve their core business is the main reason why many organizations fail to implement big data programs.

To reinforce the above theory, a research by Brynjolfsson, Hitt & Kim (2011) found out that companies incorporating a data-guided management approach, i.e., making decisions based on data and business analytics (“data-driven decision making” or DDD) achieved productivity gains that were five to six per cent higher than other factors could explain.

In order to develop and implement a robust analytics strategy, SMEs need to attract individuals with sufficient qualifications, knowledge and expertise in the analytics’ field - like Beane did when in charge of the A’s’ helm. The remainder of this report will essentially focus on the challenges that SMEs will face in their quest for big data qualified professionals.

IV. EVALUATING OPTIONS - THE FACTS

The quest will lead SMEs towards three different paths: they can hire big data specialists, they can train their in-house human resources to ensure that their staff is equipped with the right technical skills to handle big data efficiently or they can outsource the whole or part of their big data strategy.

A. Talent Acquisition

Attracting analytics professionals might be harder than it seems: SMEs that opt for this option will face significant constraints, being the most notorious a predicted shortage of analytics talent, particularly of people with deep expertise in statistics alongside managers and analysts who know how to operate companies by using insights from big data. As an example, the United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyse big data and make decisions based on their findings (McKinsey Global Institute, 2011). In the UK, a similar scenario: demand for big data staff has increased by 912 per cent between 2007 and 2012 while moderate predictions equate the generation of approximately 28,000 gross job opportunities per annum by 2017, i.e., there would be around 132,000 gross job opportunities in total created in the big data field between 2012 and 2017.

The gulf between supply and demand will have a direct impact over the levels of remuneration of big data staff: the median advertised annual salary for big data staff in 2012 was already found to be 21 per cent higher than that for general IT staff (e-skills UK (II), 2013) making this option economically less attractive for an already financially constrained sector.

B. In-house Training

SMEs should consider in-house training as an important alternative to recruitment. A recent survey by analytics company Acunu (2012 cited in Curtis 2012) - exclusively applied to members of the Big Data London group, Europe’s largest big data community - underlies this notion: more than three quarters of respondents indicated learning ‘on the job’ and ‘self-teaching’ as the most effective ways for businesses to keep up to date with the latest big data skills (including the Hadoop framework and No SQL databases – such as Cassandra and Mongo - which remain at the top of the most sought-after big data skills (Curtis, 2012; e-skills UK (II), 2013)).

Nevertheless, businesses that opt for this approach will face additional challenges as data/analytics training seems difficult to be sourced. Examples appearing to confirm this scenario are comprised in a couple of recent studies: the first by e-skills UK (I) (2013) where 45 per cent of participating businesses stated that they would be likely to have difficulty sourcing this type of training; the second by Lavastorm Analytics (2013) where nearly 64 per cent of participants stated that they couldn’t identify an external organization they could rely on for ongoing education of this nature.

SMEs need to seriously consider this option if they are looking to enlarge and consolidate their analytics’ skillset base; businesses willing to take this approach will be boosted by some positive signs including the emergence of initial training programmes around data science (Big Data Insight Group, 2012). However, as highlighted above, businesses that opt for this approach might see their efforts hampered by the existing lack of appropriate training courses (e-skills UK (I), 2013).

C. Outsourcing the Strategy

The lack of general knowledge and skills in the industry related to how to approach a big data analytics initiative will lead SMEs to this last resort: outsourcing part or even the whole of their big data strategy. A third-party consulting team specialized in big data analytics can help to accelerate this process (Lockner & Kornegay, 2011).

Although becoming increasingly attractive amongst businesses - it helps to reduce costs and to leverage technology expertise (Haynes, 2009) - this option also has its own flaws being the exchange of private - therefore sensible - data with external organizations one of the most significant. For instance, outsourcing data mining tasks involves extensive collaboration - including data sharing - across different organizations. Either raw data or the revealed information after analysis often contain business intelligence (BI) and customer privacy of primary organizations so there are obvious security concerns over the potential risk of exposing this kind of private information in this type of outsourcing activities.

Organizations that opt for this option are therefore urged to adopt, implement and make use of efficient security policies and adequate technologies otherwise their assets could be very vulnerable to security breaches (Qiu, Li & Wu, 2007). The exposure companies risk by not taking additional security steps when outsourcing their data processing to third party vendors can be illustrated by numerous privacy breach incidents such as the ChoicePoint Inc. and the LexisNexis episodes (Xiong, Chitti & Liu, 2007).

V. CONCLUSIONS AND RECOMMENDATIONS

Having a solid analytics strategy will arm SMEs with insights and capabilities once only available to corporate giants (Intuit, 2012); the report proposes three alternative routes as an optimal way to achieve that but, regardless of the chosen
path(s), SMEs will undoubtedly face different challenges and risks so they’re urged to carefully evaluate their options before taking any hasty and groundless decisions.

The demand for big data skills is rising exponentially and requires multi-level and multi-sector responses. Conscious of this fact and in an attempt to raise general awareness the UK government has called for creative partnerships between business and academia in delivering data-driven innovation and data literate graduates in order to increase the analytical workforce pool (e-skills UK (I), 2013).

Companies should encourage universities to further include analytics within their curriculums, either as common core requirements or full qualifying programs, in order to generate a substantial number of graduates with sufficient background to work as analytics specialists (Craig et al., 2012).

Employers should also raise awareness about the career prospects of analytics jobs and make them more attractive as the current unappealing aspects of these are pushing qualified talent to pursue other options.

Moneyball is about creating a competitive advantage by being innovative within a tradition-bound industry. The same applies to SMEs: if they are able to challenge the conventional wisdom by being innovative, i.e., by making good use of big data analytics, the greater will be the potential for achieving a sustainable competitive advantage over their competitors (Wolfe et al., 2007).

VI. REFERENCES


How Web Developers Can Use Trends Data To Provide Big Data Services To E-Commerce Clients

Tim Ride
University of Derby
Derby, England
100148107@unimail.derby.ac.uk

Abstract -- This report provides an overview of how trends data can benefit E-commerce developers, the paper explains how search data and trends can benefit an e-commerce developer by allowing them to provide a service to their clients through the analysis of trends relevant to the client’s products and services. The paper provides an insight into sources of trend data and how it can be collected as well as how this data can then be used to provide reports to clients that may allow them to make decisions on products, services and advertising based on current or predicted trends. Also discussed in this paper is some basic implementation decisions to be made when providing such a service and current issues with retrieving trends data in bulk.

I. INTRODUCTION
This Paper provides an overview of how Big Data is relevant to E-commerce developers, specifically those classed as an SME (small medium enterprise).

Search and trend data is becoming more freely available and this allows companies to provide big data to clients as a service or to add value to their existing services. Specifically this report applies this to web development organizations with e-commerce clients. However it is possible to leverage trends with any company that uses products and services that may appear in search results. This is achieved by collecting data from search engines on related product searches over time (known as trends). Once the data is collected it is analyzed/compared with other results in the database and used to provide a report to the client on current trends and how they relate to their specific products and keywords. When used within the context of products the client stocks, trends can also help to ascertain which product is currently the most popular and therefore allow clients to effectively price and promote these products. This paper suggests ways in which SME that develop and host web pages for e-commerce clients may be able to leverage search trends data to provide a service to clients that informs them of search trends related to their products and services.

“Search captures the interests and intents of millions of people over time, moment by moment—across every topic imaginable. In turn, the breadth and depth of that aggregated data—and its timeliness—make search an incredibly revealing and useful tool for brand marketers to examine what’s happening in their categories. Across industries—from automotive to tech to food and beverages to retail—search data can help you track and affirm existing category dynamics or shed some light on new opportunities you’re considering.” (Thinkwithgoogle.com, 2013)

From a client’s perspective this can be seen as a value added service as they will often pay for SEO and advertising to gain more clicks from search engines, using trends that relate to their products clients can advertise for the most effective keywords from their product range at a given point in time. This results in not only a higher count of customers viewing trending products but acts as a gateway to the site also. These products could also be featured on the front page of a website alongside the clients other popular selling products. This allows clients to be more efficient and plan SEO to be more effective.

II. SOURCES OF TRENDS DATA
Open data is made available by search engine providers, government, scientific studies and social networks to name a few, of these sources the most relevant to ecommerce are search engine and social data but this does not mean other data could not be used.

Search engines provide a valuable resource of trends data and Google for example freely provides this information. The data can be queried by specific terms and results in an already streamlined set of data that does not require crawling or reducing. The data can be compared with other appropriate terms and scored on which trends are currently most relevant according to the client’s products.

It may be worth considering other ways to generate trends data for example news, fashion and reviews could all provide data that may be analysed for the purpose of predicting trends. By collating and analysing this data on a large scale it may be possible to use these sources to provide an insight to clients on how trends will evolve. In other cases a client’s products or services could be related to this data, for example a client selling clothing may wish to analyse fashion trends. This approach results in more development of a system to collate the information and query it for trends.

III. USES FOR TRENDS DATA IN E-COMMERCE
Trend data is useful to e-commerce sites as they show which terms are producing the highest volume of searches on a search engine or the most talked about subjects in social media. Since most e-commerce sites rely on SEO and other search marketing for their results it would be most beneficial to use keywords that have a high trend or search volume currently.
(Vosen and Schmidt, 2011) shows that trends can correlate to a product's popularity.

Trends could be also compared by the client on multiple levels, to provide an insight on how to sell or promote specific products. An example of using trends data to compare products may be different colors of a t-shirt for example (see fig.1). Once compared stocks or promotions can be based on the product that is most popular at the current time. Another example could be comparing product ranges like trousers compared to t-shirts and deciding using this data how to provide these products. This data allows an SME web developer to provide reports on which products to prioritize and how to the ecommerce client and results in them being enabled to make decisions based on this. This data can be provided to the client in an easy to read report format each month and based on the information provided the client could decide how to change their strategy.

IV. IMPLEMENTATION OF THE SYSTEM TO GENERATE TRENDS DATA

Big Data is often associated with requiring large databases and complex reduce functions to provide results, however by accessing data that is already produced as a result of processing big data it is possible for an SME to generate a small set of relevant data to produce results from otherwise unavailable sets of big data.

The service could be provided by SME's at very little cost or risk and they may not be aware this is a viable option, however some infrastructure would need to be set up and possibly some development work to improve the system. Currently there is data available from search engine providers however there are no real API's (Application Programming Interface) for obtaining larger amounts of data. When API's are not provided for this other tools can be used or written to gather data autonomously. Using tools currently provided by other developers or developing a system to gather this information in bulk. For example Google do not currently provide an API for their trends service, however there are API's freely available to gather results on trends queries in bulk that send and collate multiple requests for example (Bertram, 2013) provides a system that can collate Google trends data in bulk this data could then be input into a database for the client.

When implementing the system it would therefore be an advantage to spend some time developing an efficient system to deal with extracting and collating large amounts of data. Once data is collated into the database graphical representations can be created for the data (as it is represented by Google trends for example) or a basic list can be provided. There is little need to have to deal with big data as such since most of this is taken care of already in this example and a more manageable database is created for each client.

When collecting data from other sources for example fashion and news or social networks a more efficient development of the system may be required. First the client would need to define what trends they are looking for and

Second, the results of this would need to be collated and the data then reduced to produce a report on how prevailing this current trend is.

An example of using other sources for data would be if a client produces clothing, the system would need to use specific trends already identified (either by the client or through a system such as Google's collate) to crawl known sources for data. This is not unlike search engines that provide a news search and this could be used to generate a data set from which to identify possible trends. While this does not accurately predict a future trend it can provide the client with an insight on possible future trends and aid decision making in relation to products they provide.

V. ASSOCIATED COSTS

As far as costing, a company could provide the service on a cloud based infrastructure and charge on a per client basis to leverage the cost of expanding this resulting in very little cost for an SME. However some development may be required and this costs the company as they need to assign a developer to work on this. Overall the system can be implemented at minimal cost depending on the aspirations at the time and how the system is implemented.

VI. CORRELATING AND PROVIDING REGIONAL DATA

As Thinkwithgoogle.com (2013) shows, Google's trends service allows searches to be filtered by locations, this provides the ability for a web development SME to provide trends by region to the client. If using a sufficiently large deployment or cloud infrastructure this may result in the ability to change search marketing and product promotions on a regional basis based on each regions search trends. If a client operates in a specific location or region the data can be filtered to provide only trends that are relevant to this. Furthermore other areas that show a high rate of searches may provide an insight on other areas to target with products and promotions.

Correlation of search terms is another important factor as finding the correct keywords to track and use in SEO can make a big difference to clients. Google correlate (a part of trends) is a tool that can map search terms over a period of time and find common correlations. It does not provide causation (the reason the terms are connected) only that there is a correlation as shown in (Lazer et al., 2014) and (Bollier, 2010) For the purposes of e-commerce causation may not be always required as if there is a correlation between products and trends data over a period it shows if they are popular or not at this specific time. An example of correlated data is shown in fig.1, here the
search term is sweater and data is show that correlates, the data
provides results such as coats as well as different searches for
sweater as searches for the two peak at the same time periods
during the winter). It is therefore possible to provide both
terms that relate to the same trend and other trends that
correlate to the current term.

Social trends may also be an advantage, using social
networks data can be gathered on popular trends and compared
with search trends to give more accurate results than on either
one alone. This may be advantageous to sites trying to predict
future trends as social networks may develop trends before
search engines. Some social networks provide API's capable of
collecting trend data and this could be used to reinforce data
retrieved from search trends.

VII. PROBLEMS WITH CURRENT RETREIVAL OF DATA

Currently there are some official API's that can be used for
retrieving trends data examples of this include Facebook and
twitter who provide API's that are capable of showing trend
data. Some providers however have no current official API for
retrieving data and this could pose future problems should they
be implemented in future. It is possible that the way in which
data is retrieved can be changed by the provider. For instance
Google's trends data can currently be accessed by changing
arguments passed to the system in a link, but as this is not an
official API it is possible that in future this could change
potentially resulting in the system having to be redeveloped.
While this could cause an issue with systems developed for
current data the use of an API would ultimately be beneficial as
developers would have a clear and supported means to access
data and therefore rewriting the system to accommodate this
would be preferable to continuing to use the current services.

VIII. FUTURE DEVELOPMENT OF TRENDS DATA AND RELATED WORK

As trends data is used more and more to provide contextual
services to users it becomes more important to provide access
to developers to access this information. Currently while the
data is available it may not be supported by the provider
through an official API, however as demand for such services
increases it may become easier for the information to be
accessed. The implementation of such an API would drastically
improve the ability to use trends data and have support
provided to the SME. Another requirement is that more search
providers could offer trends data as currently only Google offer
a trends search service.

IX. CONCLUSION

In Conclusion it is possible for an SME to leverage big data
without requiring the infrastructure setup, costs, expertise and
large databases to process this information. Using freely
available open data it is possible for a web development
company to provide a relevant and valuable service to its
clients. There are some limitations to providing such a service
and currently it is not a simple implementation as there are not
always official API's to access the data and this could result in
changes to how the data is retrieved and accessed and could
result in changing the system as the way accessing data is
changed by the provider.

While there are limitations to providing a service based on
trends data there are also benefits to adopting this strategy as
many clients spend on SEO (Search Engine Optimization) and
PPC (Pay Per Click) advertising based on keywords it would
be beneficial to provide current search trends to associate
currently relevant keywords resulting in more efficient
advertising of products.
X. REFERENCES


Big Data For Search Engine Marketing

Benefitting Search Engine Optimization For SMEs

Stephen Ridgway
Enterprise Systems
University of Derby
Derby, England
Stephen.ridgway1@unimail.derby.ac.uk

Abstract— Search engine optimisation SMEs improve the ranking of websites in major search engines such as Google and Bing. Big data is becoming increasingly important in search engine optimisation as client’s data becomes larger and data resources become available to gain information from. Open data is available freely that can be used in combination with client data to benefit the client’s website and search ranking. There are many ways to manage big data so that it is easier to store and query the information than before. Hadoop and MapReduce are being widely used at lower costs than previous methods to provide manageable big data. There are a wide selection of tools and services available that enable data to be queried to identify content and keywords to use for advertising campaigns or to improve the content of a website.

Index Terms—Search Engine Marketing, Big Data, Optimisation, Open Data, Querying.

I. INTRODUCTION

Search engine marketing is essential to businesses gaining frequent visitors to their web site whether they are a large or small business it is possible to achieve high rankings in search engines. Big data has created many opportunities to analyse and use the large amounts of data to improve search engine optimization on web pages. More open data is becoming available from Government sources, social media and many other sources that can be accessed freely. Trends on Google can be analysed and saved for future use enabling a company to build a large amount of data for their own use. Distributed file systems, programming libraries and distributed databases such as Hadoop and MapReduce have made it easier to manage and analyse big data with less resources and cost. With the right tools and experience, it is possible to make valuable predictions with data that is available and identify keywords that can improve organic and paid for search results.

II. WHY BIG DATA IS IMPORTANT TO SEARCH ENGINE MARKETING

Search engine marketing is used to increase the ranking of a web page in search engines. There are two types of results that can be brought up in the results of a search; organic and paid for results. SEO can make a website easier to crawl and improve individual pages by making them more accessible and easier to find in a search engine (Cutts, 2012). Using big data to improve content will provide more relevant data to users and increase traffic to the site from organic search results. The keywords from mining information will provide keywords that are trending and competitive.

Useful relevant content is important for optimising a web site so that it will be more searchable. Google recommends that web pages contain words that a user will search for to find the information they require (Google.com, 2014). These keywords can be identified by using big data that is relevant to a client’s business. With the huge volume of data that is available there is a great opportunity to use it to benefit clients; however the massive amount of information that has accumulated is very hard to store or analyse using conventional databases and software (Provost and Fawcett, 2013, pp. 51–59). Big data can provide information that a company cannot uncover using just their internal data (Chui and Manyika et al., 2014, pp. 102-105).

A. Predicting Trends with Big Data

Data is being made available that can be used to predict trends and improve the quality of client web sites. The application of the data can be used for marketing tasks such as targeted marketing, online advertising and the data can be used to analyse customer behavior to improve the efficiency of their content and services (Provost and Fawcett, 2013, pp. 51–59). Social media data can be used to identify keywords that can be used to bring in visitors to a web site. A study by Erik Brynjolfsson (2011) has shown that companies that use data to help make decisions have higher productivity, market value and in some cases more profitability by up to 6% (Brynjolfsson and Hitt et al., 2011). AdWords is a service by Google (also known as Pay-per-click) where keywords can be bid on so that when they are searched for they will appear at the top of the search results and by the side. By predicting which keyword is going to be popular it is possible to spend less on bids for it and gain a higher place in the search results.

B. What Data Is Available?

There are lots of sources of big data that can be queried for useful information. The types of data available to use are; open data sets, Google Trends, social networks, public data sets, government, scientific studies and client or business data that has been collected or stored. In addition, crawling the internet and collecting information about a group of websites for analysis could be useful for monitoring how well competitors are doing or how to find patterns that can benefit keyword research.
C. Google

Google Trends is a free service by Google that can be used to view how popular search terms are between a set period of time. Google correlate is a tool that can find search patterns that correspond to real-world occurrences, data can be queried or provided to view trends, when a query or data set is entered to Google correlate it uses web search activity data to find patterns, the data can then be viewed or downloaded for analysis (Google, 2014). By using these free tools, it is possible to collect and analyse large amounts of data to identify possible connections between products and services and use keywords during a time when competitors would not. Google Adwords offers users keyword prediction for their websites that can be done manually using big data sets of your choosing by using Google Trends and Google Collate (Google, 2014).

D. Open Data

Open data is information that is available for anyone to use, usually it is government data or geographical data, it has great potential to improve services if used correctly (Manyika and Chui et al., 2013). Other types of open data that are available are; social-media interactions, demographic data, financial transactions, health-care benchmarks, and real-time location data that can be used for analysis (Manyika and Chui et al., 2013). Research by the McKinsey Center for Government and McKinsey’s Business Technology Office found by using open data seven business sectors generated more than $3 trillion a year additional value, which has provided businesses establish new products and services which will improve operations effectiveness (Manyika and Chui et al., 2013).

Public data sets are collections of data that are hosted by companies such as Google and Amazon that can be queried. The Amazon Web Services public data sets are located on their cloud system which can be accessed by using their service, users pay for the computations and the storage of their own applications but the storage of the data is done by Amazon (Amazon, 2014). Public data sets solve the problem of storage as it is done by the company, this also gives you access to services such as the Amazon Elastic MapReduce (EMR). Amazon EMR uses Hadoop to distribute data across Amazon EC2 instances which are scalable to the needs of the data; analysis of data and web indexing can be achieved using Amazon EMR (Amazon Web Services, Inc., 2014).

E. Internal Data

Google crawls trillions of websites. Crawling trillions of web sites would be difficult to do, and it would need to be done frequently to provide useful data. Large amounts of data are collected in an unstructured form, to be able to query the data it must be structured so that useful information can be gained from it (Matsudaira, 2014). When using client data it is important to not use any private information that can be used to match people to the data. Zhang and Yang et al. (2013) provides a scalable multidimensional anonymization approach for providing privacy to big data using MapReduce on a cloud. Data can also be retrieved and stored by using your own web crawler or using software packages that can retrieve data, for example GNU Wget and cURL are useful for accessing and storing website content which can be queried for useful content. Once the data is extracted it is necessary to process the HTML content for the specific content that is relevant for your project then store it in a structured form that may include encoding the information as a JSON object then storing it in a database field (Matsudaira, 2014, pp. 10–11).

III. HOW TO USE THE DATA

It is problematic to manage large volume of data as it is difficult to store and process with low costs. However there are ways developed that reduce the time and difficulty such as MapReduce; which can analyse large amounts of data and Hadoop; which is a distributed file system. By using these techniques, the data will be easier to manage and analyse.

A. MapReduce

MapReduce is a programming model developed by Google to analyse large amounts of data (Khan, 2014), in which its library can be used to process tasks that use a huge amount of data much faster than previous methods. Dean and Ghemawat (2008) discuss how MapReduce can process data quickly by distributing a task among hundreds or thousands of machines. The MapReduce model was designed in a way so that it is easy to use even for programmers without experience with parallel and distributed systems. The MapReduce computation works by taking input values and outputting them, using key/value pairs. The computation uses two functions: the map and reduce functions, the map function groups together the values that share the same key and sends them to the reduce function. Then the reduce function merges the values together to form a reduced set of values. The map and reduce functions are written by the user to perform the operation required for the task (Dean and Ghemawat, 2008; Lu and Güting, 2013).

Most of the computations are run locally on the machines so that they do not consume much network bandwidth. As network bandwidth is an important resource using as little as possible is very desirable. The data is optimised and then written to a single copy to the local disk which will save network bandwidth. The user will send the program to a master machine, which will have a number of workers which the data is optimised on. The master keeps track of the progress of each worker and can control which machines are not processing correctly by checking if they respond to a ping. Due to the large amounts of machines that are used to process the data, the library must have fault tolerant procedures in place to handle the errors gracefully. If a worker machine becomes unresponsive, the master machine will know as it pings the worker machines periodically. The tasks that were in process on the worker machine are reset to the initial state, and then become available to the other worker machines to complete (Dean and Ghemawat, 2008, pp. 107–113).

B. Hadoop

Hadoop is the open-source implementation of MapReduce and is a popular tool for processing big data such as social applications, recommendation engines, unstructured data, data mining and text mining. As Hadoop processing jobs can take tens of minutes to complete, faster solutions such as Google's
Dremel service can be used for much faster results (Sato, 2012).

MapReduce and Hadoop are often used as a solution to manage big data. Borthakur (2007) describes the Hadoop Distributed File System (HDSF) as a file system designed to run on a distributed system which is highly fault tolerant and can be used on low cost systems. The applications that are typically used on HDSF are large data sets which are gigabytes to terabytes in size (Borthakur, 2007, p. 21).

C. Other Options for Querying Big Data

MapReduce is not the only way to process big data, other methods include: Partial evaluation for distributed query processing. Query preserving compression. Query answering using views, bounded incremental query answering. Top-$k$ query answering and early termination (Fan, 2013, pp. 14–28).

1) Partial Evaluation For Distributed Query Processing

Partial evaluation in distributed query evaluation involves partially evaluating the query in parallel and sending the results as compressed Boolean functions to a coordinator which joins the results to obtain the result. Buneman and Cong et al., (2006) study has shown that partial evaluation has proven useful in code optimisation, compiler generation and data flow evaluation; data flow evaluation being the most useful for querying big data as it has proven useful at performing numerical and aggregating computations over large data sets.

2) Query Preserving Compression

Fan and Li et al. (2012) research in query preserving compression for large graphs describes how it is used to convert big data to small data. Social networks often use large graphs in their applications. For example Facebook has hundreds of millions of users that are very costly to query. Indexing can be used to speed up evaluation of queries however indexing incurs extra cost. By compressing the query into smaller graphs and preserve information that is relevant to the original query it is possible to achieve a better compression ratio.

3) Query Answering Using Views

Dev.mysql.com (2014) describes Views as queries that can be stored and when invoked produce a set of results. Views are useful for producing queries quickly on large websites. Using views to query big data is not as useful as other methods as it is better for relational databases or semi structured data (Fan, Wang and Wu, n.d.). Views in NoSQL big data is not used for many cases; however it has been proven useful in querying graph data produced from social networks (Fan, Wang and Wu, n.d.).

4) Bounded Incremental Query Answering

Fan and Geerts et al. (2013) studies show that using bounded incremental query answering can be achieved on big data by creating natural classes on Boolean queries. Bounded incremental query answering does not depend on the size of the original big data but on the size of the incremental changes that have been made.

5) Top-$k$ Query Answering Early Termination

Top-$k$ query answering and early termination processes data without having to process the entire data set (Fan, 2013, pp. 14–28). Xie and Lakshmanan et al (2013) work improved the top-$k$ query answering on cached views by improving previous algorithms, introducing an indexing structure which stores cached views in a central data structure that can be used to receive results more effectively.

6) External Data Querying Options

It is not always possible to query the data on local machines however there are solutions available to be able to analyse data externally: Sato (2012) describes Google BigQuery as a public implementation of Dremel, Dremel is designed to perform SQL-like queries on large datasets. BigQuery provides data processing on large datasets for users outside of Google. The key differences between MapReduce and Dremel are: MapReduce is designed as a programming library that can batch process large datasets and Dremel is an interactive data analysis tool for large datasets. BigQuery is best used for multi-dimensional analytical queries that require results as fast as possible, whereas MapReduce is more useful for long-running batch processing. Amazon also offers many solutions for storage and querying of big data, both of public datasets that can be accessed using Amazon EC2 instances or hosted Hadoop clusters (Amazon Web Services, Inc., 2014).

IV. CONCLUSION

There is a lot of information that is freely available to use that can provide keywords and content for companies wishing to improve their search engine marketing. The increasing amounts of open data that is available to use from governments and social media alone would take a lot of resources and time to analyse, however by refining the data to a specific set it is possible to structure and analyse it using tools available such as the Amazon Elastic MapReduce cloud. I have found multiple examples of data retrieval for big data that can be used with or for specialist occurrences such as querying large graphical social data using compressed queries.

The aspects of search engine optimisation that can improve a site’s ranking can be retrieved by analysing search trends on Google and other major search engines. Clients seeking search engine optimisation often have huge amounts of data that is getting increasingly difficult to store and use, my research has found that by using distributed systems to store and analyse data it is possible to structure the unstructured data and query the data that is needed by using distributed systems. Care must be taken when using client data to ensure that client privacy is not compromised by following guidelines such as MapReduce scalable multidimensional anonymization for big data privacy that Zhang and Yang et al. (2013) recommends.

SEO SMEs have many options on how to use data they have collected; data can be managed locally or using services provided by companies such as Google and Amazon. Cloud services also provide tools to use the data such as Google's BigQuery service. The tools to manage and query data are available freely or at a lower cost depending on the situation. If large amounts of resources are needed then it would be beneficial to invest in a cloud service that offers ways to query and manage big data. However if the task is smaller, then
MapReduce and Hadoop can be set up to process the data in a relatively small environment.

V. REFERENCES


Abstract- The rate at which the quantity of data has been increasing is so high that individuals and businesses have never imagined whether such rate would be experienced. In their day-to-day activities, businesses such as Small & Medium Enterprises (SMEs) deal with several millions of kilobytes of data. These huge amounts of data contain information about their daily operations, customers, suppliers, employees, competitors, in addition to various sensors used which are commonly embedded in technological gadgets and vehicles. The high rate at which the amount of data is growing has been attributed partially to the use of social media. In addition to the storage of huge amounts of data, the provisioning of the needed storage capacity to guarantee high quality decisions from the management for timely and accurate operations is also a characteristic of big data. The paper addresses issues surrounding big data. The opportunities, challenges, benefits and operations within the emerging trend are discussed.

Key words—Big Data Opportunities, Big Data Technology, and SME.

I. INTRODUCTION

According to Sestini (2012), the term big data refers to data sets that are too complex and large that it becomes very tasking to process such data sets using conventional technologies in data processing. Activities involving the derivation of extra data from repeated analyses with a certain level of interrelation are a core reason for the existence of big data (Mann 2012). Furthermore, the unexpected growth of data sets in the current digital world also came as a result of newly developed technologies (McGuire, Manyika & Chui 2012). Among such technologies include: tools for aerial sensing, remote sensors, ubiquitous mobile devices, Radio-Frequency Identification Detectors (RFID), software logs, and IP cameras. These developments have not only made data processing and storage more complicated, it have also made various types of information more available and measureable. The greater availability of information has made it possible for businesses, large and small, to gain more insights regarding their areas of concern. However, the opportunity has an underlying cost. As a result of this growth, SMEs are currently faced by the challenge of deciding on the custodian of big data proposals in the entire organization (Jones 2014).

Ohlhorst (2012) explains that the growth of technology coupled with the rapid creation of digital information devices to utilize this technology has been a key factor in the massive increase in the amount of data being prevalent in today’s business world. SMEs as well as other businesses have a high granularity and frequency of data collection as they collect transaction data from all their clients, in addition to the storage of clients’ personal information such as the credit card numbers (Powell 2012). The high rate at which the amount of data is growing makes it necessary for additional storage capacity to be created to make sure that all the data is available in a retrievable source for future uses in decision-making. With proper identification of possible challenges, SMEs can be able to position themselves strategically to harness the benefits resulting from big data. The paper addresses the opportunities, challenges, benefits and operations of big data as used by SMEs.

Laney (2001) explains that the ability of the organization to handle huge amounts of data efficiently is a key factor in characterization of data as either big data or not. In this regard, some organizations consider data sets to be “big data” when they have used several thousands of gigabytes while other organizations conclude so after using several thousands of terabytes. The reduction in the cost of data storage as well as an increase in the speed of the data capturing and exchanges have coupled with the great increase in the number of sources of data to cause data explosion (MacInnes 2013).

II. BIG DATA: QUESTIONS OF OPPORTUNITIES

The use of big data raises alarm due to the fact that it has several sources at high volume, velocity and variety (Powell 2012). To extract realistic opportunities from this data, SMEs need to ensure that they possess the required skills, processing power and the capacity for excellent data analysis. The fast growth in the volume of data has given business people a great opportunity to gain value and a lot of knowledge that was not in existence in the past (McGuire, Manyika & Chui 2012). This form of progress has brought the SMEs to an environment with great sets of data that come along with new technologies and tools for the analysis, organization and tackling of the current and future challenges (Laney 2001).

A. Collaboration

The collaboration of employees in many organizations has been affected positively by the availability of big data. As a result, several organizations have made it a must for the IT personnel and business managers to engage in collaborative
efforts for the achievement of significant value from each set of data (McDonnell 2011).

B. Quality decisions

The use of big data by SMEs presents them with a great deal of opportunities, which includes a better understanding of how to make high quality decisions that will allow them to pursue the best action when several challenging options are present in the business operation (Mann 2012). To realize efficiencies and economies, business data, in motion or at rest, should be run close to efficient analytics. However, the management must make sure that the infrastructure has the appropriate defensive strategies to mitigate the impact of legal issues, storage risks and other related risks (Mayer-Schonberger & Cukier 2013).

The acquisition of the current skills needed for full utilization of big data allows businesses to obtain great value from this data (Jones 2014). One such potential is the ability to offer data-driven sales. A similar strategy is the application of big data to marketing efforts. This allows certain advertisements to be directed to specific commodities, which SMEs can use to gain new revenue prospects and engender innovation (Powell 2012). The fact that the big data and its related analytics can be hosted over a cloud suggests that cloud computing can be used to reduce the pressure experienced by the IT infrastructure of SMEs. As a result, the SMEs are able to achieve the scalability, flexibility, expandability and economy necessary to gain competitive advantages (Mann 2012; McDonnell 2011).

If big data is managed effectively, economies can be transformed through the creation of growth in productivity, successful marketing, and consumer surplus. SMEs can also obtain competitive advantage at the marketplace through exploitation of extreme analytics and reports that allow managers to make correct decisions when dealing with important transactions, such as supplies and purchases (McGuire, Manyika & Chui 2012).

III. BIG DATA: CHALLENGES

Despite the many opportunities and value offered by big data, it also presents significant challenges. For SMEs to enjoy benefits offered by big data, they must be able to safeguard the privacy, governance and security that will ensure that all data and insights are protected. It is worth noting that the use of big data is a great threat to the privacy of business data. The use of big data challenges businesses as it tries to acquire, store, share, search, transfer, visualize and analyze data. This is due to the fact that businesses often find it difficult to tolerate the long period of time required to complete the tasks (McAfee & Brynjolfsson 2012).

As explained by Laney (2001), it has become a major challenge to analyze trends due to the current social wave. Most of the customers are in social activities and blogs as well as other online activities. The SMEs are required to aim at pursuing the big data from such arenas to avoid being outdone by their competitors who are known to take advantage of such channels for the utilization of the same data (McGuire, Manyika & Chui 2012).

A. Legal Challenges

Several legal issues have been brought about by the increased economic importance derived from big data. Those involved in copying of big data, its usage and combination with other types of data expose themselves to several legal issues that become more evident when several businesses are involved. The challenge manifests in the event of determining the owner of a specific set of data and when there is a need to monitor the manner in which the big data is used (Mayer-Schonberger & Cukier 2013). The businesses are also finding it a big challenge to determine the person responsible for any negative effect caused by the use of a certain set of data. The SMEs are therefore required to address all the legal arising in their use of sets of big data if they want to derive great value from this data (McGuire, Manyika, & Chui 2012).

B. Technological Challenge

To mine the benefits that lie in the use of big sets of data, the SMEs are required to ensure that there is implementation of current techniques and technologies including data storage devices, analytical applications and big data policies. As far as the data maturity of SMEs is concerned, a number of technology issues are experienced. There is an overhaul needed by conventional systems, which are mainly found to be incompatible with the newly emerging trends in the sets of data (Peck 2013; McGuire, Manyika & Chui 2012).

The format incompatibilities being experienced hinder proper integration of data in addition to preventing sophisticated data analytics making it difficult for the SMEs to enjoy the potential benefits of big data (Simons 2013). It is important to note at this point that innovations and the invention of technologies and techniques are essential for the exploitation of the potential hidden in big data. The need to carry out continuous innovation to allow the SMEs to integrate, visualize, analyze and consume the data, which is expanding at a very high rate, has brought more challenges due to the huge amounts of financial resources required (Ohlhorst 2012; Peck 2013).

Many of the SMEs have no expertise and tools to manage the huge amount of big data and the high velocity at which this data is generated. Despite the cheap means of data storage, the lack of proper structure and huge variety of data brings a huge challenge in parsing formatting, scrapping and transforming big data (Powell 2012; McGuire, Manyika & Chui 2012).

C. Organizational Challenges

Another factor of great importance is the initiative to deal with lack of proper understanding on big data management within SMEs. There is also the need to deal with the lack of appropriate mechanisms for the customization of applications in order to realize the full potential of big data. The other challenge to deal with is the absence of appropriate ways for the realization of the vision of the SME and capitalization of optimal capacity of big data to attain competitive advantage over their competitors through establishment of new market segments (McGuire, Manyika, & Chui 2012).

In a bid to ensure that the infrastructure is dynamically responding to business processes, the infrastructure must make
use of real-time information circulating within the business. If this is adhered to, SMEs are known to achieve a higher level of agility and improved economies. To realize efficiencies and economies, business data, in motion or at rest, should be run close to efficient analytics (McGuire, Manyika & Chui 2012). However, the management must make sure that the infrastructure has the appropriate defensive strategies to reduce the legal issues, storage risks and other risks (Mayer-Schonberger, & Cukier 2013).

IV. BIG DATA: BENEFITS

There are several benefits obtained by SMEs from their use of big data since there is presentation of transformational prospects for the purpose of value extraction (McGuire, Manyika & Chui 2012). Through the exploitation of big data, both stored and in motion, SMEs are offered a platform they can enjoy with several benefits such as business survival and financial gain.

A. Survival In The Market

The analyses of big data has allowed SMEs to create advanced market capabilities and are thus more able to exploit the market in a better way than when using conventional data analysis tools (Maclnnes 2013). Conveniences in terms of access, processing and analysis has given the SMEs a chance to do better analysis of customer and supplier trends, in addition to the proper comparison between the company’s products and its competitors (McGuire, Manyika & Chui 2012). SMEs have also achieved a faster searching and processing of data through provision of big data to the various parts of the organization. Furthermore, the integration of big data in a company gives room for concurrent operations, which improves marketing time and the quality of operations (Peck 2013).

If big data is managed effectively, economies can be transformed by fostering growth in productivity, successful marketing, and consumer surplus. SMEs can also obtain competitive advantages at the market by taking advantage of extreme analytics and reports. These resources allow managers to make better decisions when dealing with important transactions, such as supplies and purchases (McGuire, Manyika, & Chui 2012).

B. Financial Gain

After ensuring business survival, it is a natural goal for companies to pursue financial gain. SMEs profit from the utilization of big data. The use of big data has given SMEs a chance to generate high efficiency and high financial value (Powell 2012). As an example, a 60% increase in the possible net margin has been experienced in the United States while an annual efficiency growth of about 0.75% has been experienced in the same businesses. There has been positive reception of the benefits of big data use, which has prompted SMEs as well as other businesses, to invest highly on technology in order to obtain financial gain from big data (Powell 2012).

V. BIG DATA: OPERATIONS

A. Appropriate Personnel

Laney (2001) asserts that the implementation of a valuable and fully functional big data requires SMEs to set the appropriate personnel who can formulate the steps and identify critical factors towards this. It is worth noting that technology is not enough to utilize big data effectively. To create a big data infrastructure, SMEs should ensure that they implement important technologies for the support of data infrastructures and IT tools required in platforms of big data. There is also a need to collect both external and internal data to increase the domain of the data set accompanied by acquiring of big data expertise and talent (McGuire, Manyika, & Chui 2012).

B. Correct Technologies

Simons (2013) argues that there are two technologies associated with big data. There exist applications responsible for interactive and real-time workloads with primary function of capturing and storage of data. The other applications provide analytical capabilities for analysis of complex data. Deployment of these categories as a single application is common since they are complementary. SMEs can either self-host technologies associated with big data or seek services from cloud providers. Cloud is preferred most of the times as it provides cheap software and hardware platforms as well as elastic, flexible, and on-demand services (Laney 2001).

VI. CONCLUSION

The growth of big data provides significant opportunities capable of increasing the amount of data as well as the sources and the speed of incoming and out coming data. SME managers are advised to commit themselves to the clear understanding of the opportunities presented by the use of big data as well as the threats associated with its use. This should be followed by assessment of several elements for the alignment of data strategy to the appropriate mechanisms. However, SMEs face various legal, organizational, and technological challenges along with the implementation of the technology. The use of big data makes it necessary for SMEs to use complex analytics for risk reduction, better decision-making, and quick recovery in the event of threats. This has been the biggest challenge, threatening successful adoption of big data for SMEs. The capacity of SMEs to advance on this area in the utilization of big data would allow them to realize the benefits of the technology.

The risks associated with the use of big data should be mitigated. Therefore, apart from the technologies dealing with the use of big data, SMEs should also employ defensive measures in order to shield them from loss. Furthermore, more research should be done in this area in order to fully utilize the potential of big data in improving business processes.

REFERENCES


Big Data For SMEs: Questions Of Opportunities, Challenges, Benefits And Operations.

Quantum Computing

Tomas Satala
Faculty of Business, Law and Computing
University of Derby
Derby, England
T.Satala1@unimail.derby.ac.uk

Abstract—This document focuses on critical evaluation of Quantum Computing and how it can be used to deploy advanced computing mechanism and power it harnesses to assist Small and Medium Size businesses to help them with issues associated with Big Data. This document provides users with detailed descriptions of the technology and how it works, how it can be deployed, managed and used accordingly to help business with big data analytics and overcome main characteristics of big data such as Variety, Volume, etc.

Index Terms—Data, Quantum, Variety, Volume, Internet of Things.

I. INTRODUCTION

The need for new era of computing power and capability is necessity in order for the world to sustain and manage the ever-increasing rate of data produced by consumers, services and devices around the world (Barnatt, 2014). The need for change is important as we are now getting close to limits of current computing technology. That is, we are hitting the limits of processing capability due to the nature of the technology involved in development of computer processors (Jammy, 2010).

Big data is commonly referred to as an exponential grown of data. Simple definition of why we refer to it as a “Big Data” is because it allows us to generate value from a rather very large data sets that often cannot be analysed using the traditional computing technologies (Arthur, 2013).

II. BIG DATA

Many of today’s available technologies and services such as online shopping, social networking, advertising, computer cookies, Internet of Things, loyalty cards, passive data such as location based information of user’s smartphone, etc. These technologies/services are often capable of generating data that gets stored on company servers, but due to large quantities of it, they often cannot be all processed accordingly in order to deliver the benefit the business desires (Sagiroglu and Sinanc, 2013, pp. 42–47). For example, up until recently, the loyalty scheme cards were collecting enormous sets of information and data, which would just sit on supermarkets servers but would not be processed at all (Barnatt, 2014). Inside those large data sets, which would often hide very valuable information that businesses would use to generate value from (White, 2011). That is, for example, supermarkets now use the loyalty cards to monitor users purchase activities and find out what their most favourite items they buy are and therefore allowing them to send targeted offers and promotions attracting customer’s attention.

Reasons behind the big data are simple; retailers are building databases of recorded customer’s activity. Organisations from logistics, finance, etc. are also capturing more and more data. Social media is creating vast numbers of digital material that often unstructured, such as images and videos (Scarfi, 2012). Many are trying to work out alternative ways of extracting meaningful information from unstructured data such as the images and videos and recent developments from Optical Image Recognition now allow such features that permit extraction of data and information from still images and video (Uchida, 2013, pp. 523–549). For example, Google has invested lots of time and effort to deliver their image scanning algorithms services. Google image search now allow users to simply drag and drop an image into the Google search bar, which is then sent to google servers for further optical recognition and users is presented with images that are similar by colour, pixels, resolution, text, etc. (Zhou and Chen et al., 2012, pp. 98–101).

Internet of things is also on the raise with more objects becoming digitally assigned tags allowing easier communication across devices. Now more devices not just in retail but also at home are being tagged using RFID technology and slowly becoming another factor, which generates big data (O’leary, 2013, pp. 53–65).

In science, Big Data is starting to play an important role as well. However, it is worthless as the scale of the data sets is enormous and most of the data is ending up unprocessed leading to loss of potentially valuable information (Zikopoulos, 2012).

A. Big Data Characteristics

1) Volume

Volume often has characteristics of greater challenges but also delivers the most opportunities. With large scale of data, business and organisations benefit from it as it helps them to understand their customers better and guides them with a better
and more efficient allocation of their resources. For example, if people tend to purchase more of a particular item compared to other similar yet the taste is different, organisations can allocate more resource into making this particular item of taste, as it is more desired (Wu and Zhu et al., 2014, pp. 97–107).

The major down side to volume of the data is that the current relational databases have not been designed to handle such scale of data (Moniruzzaman and Hossain, 2013). They often cannot handle the flow and storage thus only leaving businesses with one option, which is to get rid of the data at random, if it has not been structured.

2) Velocity

This characteristic of big data is often known as the rate of data indigestion. That is, at what rate can your IT system intake information in order to make rapid decisions (Dumbill, 2012). However, due to large number of data flowing, the speed at which computing system need to digest the data is often not sufficient leading to lots of data ending up thrown away during disposal or completely missed. The current computing equipment deployed in organisations are often using IT system that cannot handle the rate of the data flow. The issue is that more consumers desire more data to be accessible in real time such as video streaming such as Netflix. Now with more than ever portable devices such as tablets, users often desire high quality content streaming in real time. This often puts a heavy load on current system infrastructure (Barnatt, 2014).

3) Variety

The variety of data type to be processed is becoming increasingly diverse. That is, organisations are no longer just collecting traditional data such as text documents, numbers, etc. (Price and Flach, 2013, pp. 569–574). Now more modern data types are finding their way into organisations such as Photos, Videos, Audio, Location Data, etc. For example, location based data also known as passive is when a smartphone reports its estimate physical location to servers at a certain of the time (Chen, 2011). Sometimes it can be every hour. Most of the data is unstructured and therefore it is much more difficult to process or organize.

III. EVALUATION OF BIG DATA AND VS

Due to issues and challenges posed by the Big Data’s Vs, organisations have no choice but to collect or ignore the large quantities of potentially valuable information. The current IT Technology system are simply not up for the task to handle the volume of the data. Therefore, most of the data is not processed and valuable data is often lost during the disposal.

A. Current trends in Big Data Technologies

Many large organisations such as Microsoft and Google are now realising the issues raised by the Big Data Vs, therefore deciding to invest into the future development of better computing facilities, understanding the importance and potential of Big Data (Microsoft, 2014). This study argues that with enough attention and awareness, further developments in computing technologies will allow better data collection, processing and disposal without losing the valuable information inside the Big Data (Pitt and Bourazeri et al., 2013).

1) Hadoop

Currently, amongst many organisations, Hadoop is a popular technology in Big Data, which allows faster and more efficient data processing. Hadoop is an open source product by Apache. This tool was first in its class providing scalable distributed computing system capable of handling large data (Hadoop, 2014). For example, LinkedIn is amongst many, which use this implementation. LinkedIn uses this technology to analyse personal accounts with related characteristics providing 100 Billion personalisation per week (Wong, 2013). For example, giving suggestions to user about other people you may know. Even Facebook is now deploying similar technology where they can provide users with more personalised experience allowing them to suggest people other users may now.

a) Hadoop Technology Overview

Hadoop features two components of its own. First component is known as Hadoop Distributed File System (HDFS). This File system is more structured and scalable allowing high bandwidth, cluster based storage (Hadoop, 2014). The second component is an extension of Googles own data processing framework called MapReduce. This component works by distributing large data sets in parallel processing dividing them into smaller sets and across multiple servers at the same time. This allows data processing that is more efficient. When data is processed, it is then summarised thus allowing pre-processing of the data set before more detailed analysis tools are applied (Dean and Ghemawat, 2014).

2) Cloud Based Data Processing

Most of the organisations these days often tend to use their old technology and use Hadoops implementation on top of that for data mining and warehousing. However, thanks to explosion of recent cloud technologies, from companies such as Google, Amazon, etc. organisations no longer need to have their own implementation of data analytics, as their data can be directly stored and hosted on the cloud. This allows organisation to use cloud based analytics tools and big data processing technologies. This is also rather efficient, as organisation does not need to download entire data set onto their local servers before they can start analysing it. Amazon is amongst many of organisations providing such Cloud based data analytics. They are known for hosting some of the largest data sets such as US and Japans Census Data (Eastman, 2008).

B. Organisations Investments into Technology

Big Data processing is slowly improving as more organisations are investing and producing awareness campaigns to make more businesses, mainly small one to buy into the cloud and Big Data Technologies.

IBM is one of the big names that believe that Big Data is a major opportunity for finding new and emerging data types, allowing business to run more flexibly and efficiently (IBM, 2014). O Reilly suggests that Big Data can potentially improve analytical insight as well as allowing for development of new products and services (Dumbill, 2014). Others such as Google
and Facebook have already implemented Big Data very well (Garlasu and S et al., 2013, pp. 1–4). For example, as mentioned earlier, they use data collected for detailed analysis and use it of some of their services such as their People You May know service, where you are provided with suggestion of people that you may know because of the interests, location, and other factors that you share amongst other people (Facebook Help, 2014).

Twitter also realized the importance of the Big Data and they are now using data processing on their 12 terabyte of tweet written per day. This allows them to create summary of what people feel and have been saying in the past day. For example, trends in Olympics, etc. This is also known as Current Trends or “Trending” on Twitter (Wired UK, 2014).

McKinsky Global Institute recently published report on explaining that US could save on healthcare services by deploying more efficient Big Data analytics and processing technologies. They suggest that USA Healthcare could save up to $300 billion on efficiency and quality of their services per year. They further suggest that $149 billion could be saved in administration of government across Europe. USA has recently invested $200 million in improving the tools and techniques for big data projects (Manyika and Chui et al., 2011, pp. 6 - 7).

IV. QUANTUM COMPUTING

The silicon, which is currently used for manufacturing microprocessors, is almost at its peak due to its physical limits leading to unstable functionality. Miniaturisation of transistors on Intel Microchips is currently at 22nm manufacturing process. The dead end for silicon-based chips is expected to occur around ten years’ time (Ward, 2011). This is why research into computing technology such as Quantum Computing is on raise to bring theory into practice allowing more intelligent and efficient data analytics.

A. What is Quantum Computing

Quantum computing is the next big thing in computing. It goes beyond what you might have heard or are aware of because most of the theories are an emerging science reaching out above the laws of physics. Because of all the large proportions of data created each second, there is currently no viable solution to process all data, leading to significant loss for organisations and businesses. (Ross and Oskin, 2008)

To understand Quantum computing, it is essential to know some of the background knowledge related to how current computer microprocessors work. In computer, we have a Central Processing Unit (CPU), which consist of transistors. Transistors can either be turned on or switched off. This mechanics is used to represents value of “1” or “0” (Jamal, 2011, pp. 229–232). In quantum computing however, the chip is based on Quantum Mechanical states instead of transistors. Computers use the values of ones and zero to store and process binary bits. In quantum computing, qubits are used to represent the data. In theory, quantum chip can use spinning direction of an atom to represent each qubit. Even more, there are concepts that propose the use design with rate metals and their spin of magnetic field to represent the state of position (Beth, 2000, pp. 735–736).

Because of abnormalities in laws of quantum physics, each qubit can represent values “1”, “0” just like in traditional bits, but also both numbers simultaneously. The reason why this is possible is that the atomic particles that are used to represent qubits can exist in more than one state at exact same point in time. In Quantum computing, we refer to this state as a superposition. With this capability, single qubit has ability to process a broad range of values allowing much faster data processing than conventional computer (Buhrman, 2000, pp. 131–141).

B. Overcoming Big Data Vs’ with Quantum Computing

Due to powerful parallel processing power of Quantum Computer, it can be used for a very efficient processing of tasks such as artificial intelligence, optical recognition for images or videos, scientific research data, etc. The true power behind theories of quantum computing and quantum physics allow for complex algorithms development, which go beyond traditional computers used today.

As data produced by services and devices is growing rapidly each year, quantum computers can essential help with processing of the vast data sets at real time speed when compared to traditional implementation such as Hadoop.

Quantum computers will allow all of the generated data to be processed, allowing businesses to generate value from before unprocessed data. Google and NASA are amongst two organisations that have already started implementation of D-Wave One Quantum Computer built by D-Wave Company based in Canada (Hsu, 2014).

With Quantum Computers, all data, not just some, will be processed due to incredible power of parallel processing within the quantum computers. This will increase businesses to have a lot better insight into their customers and better understanding of what people want or need.

Velocity will be no issue with parallel processing power that quantum computer harnesses. To get a better understanding of how powerful and fast Quantum Computer in theory is, let us have a look at how current computers are used in security sector in banks for example. Currently, banks rely heavily on some of the supercomputers today to generate a prime number to create secure encryption keys for customer’s bank account when performing online banking (Bowyer, 2009).

With Quantum Computer, banks will be amongst the first customers to acquire such power in order to create a far robust encryption as current encryption keys can be decrypted within moments by using complex parallel processing of quantum when compared to traditional computing system deployed today (Curtis, 2014).

David Meyer, mathematics professor at University of California in San Diego has collaborated with a graduate student in physics, Tom Wong and have proposed an algorithm only possible with Quantum Computer. This algorithm allows processing of unstructured data allowing us to overcome today issue of Variety of data characteristics in Big Data field. It is based on Einstein condensate allowing use of superposition of a particle like photon and then cancelling out the incorrect answers and yielding a single correct state. This in theory can bring many benefits of speed of quantum computer into Big
Data analytics as the time it would take a traditional computer can take an exponential time when compared to traditional computer. The actual figures currently cannot be measured due to odd nature of superposition of atoms (Meyer and Wong, 2013, p. 063014).

V. CONCLUSION
With enough research and development, we could be soon or later carrying computers based on Quantum Computing Physics. Exponential growth of data is forcing organisations to look for alternatives in computing technology in order to keep up with the demand of consumer products and electronics. Is the quantum computing the answer to Big Data Vs? Thomas J. Watson from IBM research centre has developed a model, which provides some evidence that the same level of computational power achieved by D-Waves One Quantum Computer can be solved by computer based on traditional computing mechanics (Shin and Smith et al., 2014). Seeing companies such as Google, NSA and others investing heavily into D-waves technology can prove T.J Watsons wrong, as more evidence that is robust needs to be provided to prove the Quantum Theories mistaken. For the future, businesses and organisations, whether small or big should certainly keep an eye of what is to come next as some breakthroughs are to take place, changing how businesses benefit from Big Data generating bigger value than ever.

VI. REFERENCES


The Power Of Big Data In Retail Industry
Analytics For Retailers

Umar Shakil
School of Computing and Mathematics
University of Derby
Derbyshire, United Kingdom
u.shakil1@unimail.derby.ac.uk

Abstract—The aim of this report is to critically evaluate and discuss the importance Big Data in the retail Industry. It will firstly highlight the expectations of retailers from big data and the major challenges they are facing in using it. Furthermore, significant light will be shed upon the usage of this data collected through smartphones, websites and social networks: its benefits and privacy concerns of users. Retailers will be instructed on how to improve the collection and management of big data and how, by doing this, they can achieve huge benefits in their businesses.

Index Terms—Passive Data, Smartphones, Small And Medium Enterprises, Big Data Analytics, Retail.

I. INTRODUCTION

The term “Big Data” refers to enterprises and technologies that involve data which is massive, too diverse, or fast-changing for the conventional technologies to be addressed efficiently. Though, with the new technologies, it is possible to realise value from Big Data. For instance, retailers can track web-users clicks to identify behavioural trends in order to improve their campaigns, stock and pricing. On a bigger scale, Governments can detect and track the emergence of disease outbreaks via social media. In other words, “Big Data” describes large and complex data sets that are impractical to manage with traditional software tools (Mongodb, 2014).

It is expected that the big data is going to become one of the key factors of competition for growth of individual firms. This is why all companies need to take big data seriously. In most industries, established competitors as well new entrants are expected to leverage data driven strategies in order to innovate, compete and capture values from real time information (Maniyyika et al. 2011).

This report will aim to discuss the effectiveness of big data to businesses. Particularly focus will be shed upon the benefits to the retail industry through passive data collection from consumers’ smart devices.

Most of the world leading companies in the retail arena, are now seeking to collect passive data through websites and especially smartphones. Smartphones are the most recent advance in the mobile phones category. These powerful devices offer advanced computing and communicational capabilities such as, for example, internet access and geo-positioning services. Smart devices have achieved a pervasive presence over the last few years and has important implications for society, and holds a lot of potential in the retail industry (Boulos et al., 2011).

Passive data collection is the process of collecting opt-in data from smartphones devices. The data is collected through a mobile application that runs in the background, with the user’s consent. The information gathered allows you to understand how smartphones are being used from their users and their behaviours (SurveySwipe, 2013).

Fig. 1. Types of Data Marketers Worldwide would like to Add to Their Customer Data Profile (Precisionmatch.net, 2014).

In other words, retailers, with the use of big data, want to know how much money they can make from a certain customer according to where the purchase is being made, and what is being purchased (Houston, 2014).
Although big data is very beneficial to businesses in general and Retail in particular; it still faces numerous challenges to be used widely by companies. This is mainly because marketers continue to focus on data collected through customers’ surveys and transactions. It is a hard to believe fact that even nowadays, not many businesses are collecting information from online activities, such as mobile computing and social media. Only 19% collect customers’ data from mobile channels and 35% of the business collected information from social media (MarketingProfs, 2012).

The graph below, shows the biggest challenges to use of big data for marketing according to the personal of numerous brands.

![Graph showing biggest challenges to use of “big data” for marketing](image)

Fig. 2. Biggest challenges to use of “big data” for marketing (CMTO, 2013).

The graph shows some rather interesting facts. It is interesting to note that among surveyed senior marketers, 51% say that the lack of data being shared across the organisation is preventing them to measure their ROI (Return on Investment), while 41% believe that the data available to them is not being used properly to effectively personalise their marketing communications.

Furthermore, the graph also indicates how the data is being collected too infrequently and moreover not in real time. Retailers must comprehend quickly how useful the big data can be to their business and run strategies that aim to collect real time data from their consumers.

### III. EVALUATION OF BIG DATA IN RETAIL

Big Data is not just about the size of the data, it also includes the variety and velocity of the data. Volume is a relative word for the reason that some companies may have mere gigabytes of data storage as opposed to some big global enterprises that might petabytes of data to store. The data can come from a variety of sources in a variety of types. With the explosion of sensors, social networking as well as smart devices, data within an enterprise has become very complex to semi-structured as well as unstructured data. The velocity of data in terms of the frequency of its generation and delivery is also one of the main characteristics of big data. In the field of big data, velocity must also apply to data in motion: the speed at which the data is flowing (iDA, 2012).

According to Hemsley (2014), there is such a huge amount of data about consumers available on web that the marketers will be forced to pay attention to this phenomena. He further adds that the big data is next frontier for innovation, productivity and competitiveness, however, many companies are still missing out the opportunities because they lack expertise in the data management field.

With Big Data it is possible for companies to better understand customer’s behaviours by analysing their online browsing and searching histories. For instance, companies can learn alternatives to a certain product that customers look at when considering buy a product and how they put together their shopping baskets. Amazon is a good example in this area. The website uses a technology called collaborative filtering technology, it allows to generate automatic recommendations for customers based upon their purchase history (Hemsley, 2014).

A good understanding of the consumers’ behaviour, can help companies identify valuable up-selling and cross-selling opportunities. However, it is important for businesses to identify information that is relevant and to do so companies must hire, develop and retain skilled analysts who are capable of distinguishing relevant from irrelevant data (Breuer, et al. (2013).

Collection of passive data, through smart phones, can be beneficial in the retail industry. Collecting data from these devices is very useful particularly because they are with the users at all times (Heggestuen, 2013). Passive information collected from these devices such as Wi-Fi data or Bluetooth data in order to monitor consumers’ movement throughout the stores. The information obtained can be used to understand locations that are more or less visited by the customers and thereafter an improved display policy could be applied to create revenue optimised floor plan (Rijmenam, 2013).

Walmart, the biggest retail company in the world, tracks consumers online to understand their behaviour and improve their campaigns. The company, has mainly targeted mobile technologies to track consumers and market products. This is because Walmart states that’s more than half of its customers are smartphone owners. Customers, who carry smart devices, are of particular interest to the business of Walmart, as the company estimates that these consumers spend 77% more than other customers every month and make four more trips in comparison (Walmart, 2013).

Gibu Thomas, Walter’s global head of Mobile, believes that the future of retail consists in a personalised experience for each customer delivered through the smartphone (Thomas, 2014).

In order to successfully manage, handle and deal with Big Data, retailers first fully need to comprehend all the V’s composing the big data.

#### A. Volume

Retailers must be able to process large volume of data and its management is critical to the success in global market. This is because, it is estimated that 2.5 quintillion bytes of data is created every day. Retailers must invest in competent employees who can identify and analyse relevant data from other data which is not relevant, so that they can achieve great marketing advantages (Biehn, 2013).
B. Variety

In the past, almost all the data that was created was structured data, which was very easy to manage and analyse. Nowadays, though, it is estimated that the 90% of the data generated is unstructured. In fact, data comes in so many different formats today: structured data, unstructured data, semi-structured data or even complex structured data. Different types of data, requires different type of analysis and tools in order to be analysed. For example, Tweets or Facebook posts about your brand can help retailers to get an insight of a public view of their brand while apps can generate data about what items/accessories are being visited more often and how that brand can improve their campaign as well as pricing (Biehn, 2013).

C. Velocity

The velocity is the speed at which the data is being created, stored, analysed and subsequently visualised. In the big data field, the data is created in real. With the wide diffusion of internet connected devices, machines can now pass-on real time data at the moment of their creation. The main challenge for retailers at this point is to cope with the enormous speed at which the data is being created (Biehn, 2013).

The three V’s listed above are the traditional V’s that are important to big data, however, Rijnenam argues that the 3V’s are not sufficient to describe big data and there are some other V’s that require the attention of organisations (Rijnenam, 2013), in our case retailers, in order to develop a big data strategy.

D. Value

Big data means big business and in the coming future every industry will benefit from it. It is estimated that the annual value of big data to the US Health Care is $ 300 billion. Obviously, data as it is has no value. Data becomes valuable when it is successfully analysed and the information is converted into knowledge. The value, in this particular scenario of retail industry is how, retailers use this data for their market strategies and decision making (Biehn, 2013).

E. Veracity

It is completely worthless to have data coming at a high speed if it is incorrect because it can cause a lot of problems for retailers as well as the consumers. This is why, retailers need to make sure that the data is correct as well as the analysis being done on that data. This becomes even more important in automated decision making, where no human supervision is involved. It is expected that by 2015, 80% of all the data will be uncertain as the number of networked devices and social media accounts will be double the entire population and therefore the data will be highly uncertain in both its content and expression.

F. Variability

Big Data can be extremely variable. This is a very difficult aspect of big data because words have different meaning when used in different contexts. This aspect of the big data could be used to perform sentiment analyses. In order to do so, retailers must develop algorithms capable of firstly understanding the context and then be able to decode the exact meaning of the word in the context. Again, this can be a very difficult procedure to perform and highly qualified staff is required to execute it (Rijnenam, 2013).

G. Visualisation

Arguably the hardest part of big data. Making all that volume of data comprehensible in a manner that is understandable and easy to read. Raw data is essentially unless it is analysed and visualised which means complex graphs that may include many variables while still remain readable and understandable. It is expected that in the future, visualisation will be the difference in the retail industry as it will help retailers answer questions they did not know to ask (Rijnenam, 2013).

H. Validity

Data must be validated in order to achieve accurate results. The validity of the data collected and the subsequent analysis performed on it must be accurate if it is to be used for decision making process by the retailers. For example, Businesses in retail industry could use twitter streams about their brands in combination with their website traffic to determine what the major trends in market are as far as their products are concerned (Hurwitz, 2014).
I. Volatility

An important challenge for retailers is to establish for how long they need to keep the data gathered. With some big data sources, it is only necessary to gather the data for a quick analysis after which they can be deleted from the system. This process is very important especially when the storage is limited so it ensures the rapid retrieval of data (Hurwitz, 2014).

Big data can provide insights into the types of interactions customers would like to have with the sellers as well as telling them if customers are having the desired experience with the company and how this could be improved. Even though, we are just at the start of the big data revolution, these kind of insights can be available to even retailers of small calibre via the cloud and big data. It can provide big data analytic on demand, which will make a very highly personalised and data-centric shopping experience a norm in a not very distant future (Viralheat, 2013).

IV. CONCLUSION

Use of big data can bring vast benefits to the retail industry and assist its growth by reinforcing the marketing strategies. Consider that, your store is stocked with exactly the right products that the customers want to buy and that too with the precise quantity. The business’s website contains the adequate key words which makes it easy to be accessible by more consumers (Laws, 2014). All this, is more than possible with the use of big data. Retailers must employ staff that is capable of collecting, managing and analysing relevant data in order to benefit from the huge benefits that arise with the use of big data. Since global data quantities are expected to rise every year by 40% and IT spending is expected to increase by only 5%, the firms that will invest more on big data, are more likely to reap the benefits (Hemsley, 2014).

V. REFERENCES


Can Big Data Help SMEs Especially In Theatre?

Diana Silva Caires
Faculty of Business, Computing and Law
University Of Derby
Derby, United Kingdom
100176764
D.Silva-Caires1@unimail.derby.ac.uk

Abstract— For years the world of theatre has been insecure about its future. Fortunately, the solution could lie in the opportunities of using big data. It can increase the amount of people who see performances, giving UK theatres a much needed boost, allowing it to advance and compete with theatres worldwide. In spite of this, there are challenges to consider when using big data, which raise several issues and important questions. This report will look at the multiple V’s of big data, exploring the opportunities and challenges that each one brings, in order to see if the investment in big data is worth it.

Index Terms—Big data, Theatres, Vs model, Opportunity, Challenge.

I. INTRODUCTION

Big data is defined to illustrate the expansion of available data processed in order to produce accurate analyses and facilitate businesses with decision making (SAS, 2012). It has a variety of V models including the main 3Vs that defines big data which are volume, velocity and variety. Theatres are one of the oldest form and central part of arts and cultural in the world (Mermiri, 2013). Over a two year period, Britain's theatre audiences has fallen by 9% from 14.1 million to 13 million (Telegraph, 2012). According to the BBC (2013) it states that half of small theatres' venues are at risk in London. For the arts and cultural sector, big data can bring an area of novelty to engage with (Digital R&D, 2012) and consider financial and operational failures which could embark on utilising the data better (Lilley & Moore, 2013).

With the era of big data now being used in multiple corporate enterprises, now is the time for small and medium enterprises (SMEs) to exploit this opportunity. One area that could exploit this is the theatre industry. It can be used to increase audience numbers and deliver great performances. There are businesses which are having concerns with comprehending the technology surrounding big data (Passingham, 2013). A study has estimated less than 0.2% of UK SMEs are taking advantage of big data analytics (Passingham, 2013). Most SMEs in the UK are not making any use of big data (Passingham, 2013) as if SMEs sees big data as a failure they will take no notice of the big opportunity which big data offers (TM Forum, 2012). In order to examine the opportunities/benefits and challenges/risks for theatre using the characteristics of big data, this article will explore big data Vs model individually on how it can improve SMEs theatres.

II. THE OPPORTUNITIES/BENEFITS OF USING BIG DATA FOR THEATRES

There are many opportunities among the trillions of bytes of structured and unstructured big data. However, the sheer volume needs to be narrowed by people's creativity, in order to find hidden value from the accessible data (McDonnell, 2011). By utilizing this creativity, theatres (part of the arts and cultural sector which is in essence of a creative sector) can easily think new ways in storing and analysing data, providing a great opportunity for using big data (Liebenson, 2012).

A. Volume

Today's digital age, large volumes of data are being produced every day. Over the internet more data are being generated every second than it was 20 years ago, in 2012 about 2.5 exabytes of data were created and it's been doubling as time progresses (McAfee & Brynjolfsson, 2012). For instance, TicketMaster (a popular online ticket purchasing in the UK and worldwide) collects data from 60 million individual customers across 12 markets (including theatre performances) in order to understand the audiences (Mermiri, 2013). This brings an opportunity for theatres to collect a mass volume of data to improve shows and understand their audience’s theatre interests. Furthermore, it gives an opportunity for increasing employment in theatres having analytical people to arrange and analyse large amount of unstructured data for the business (Harbert, 2013).

Nowadays, the digital age of social media from Twitter to blogosphere are being used by theatres (Gardner, 2010). From social media sites theatres can receive massive amount of data which can reveal important information about the audiences (Smith, 2014). For instance, there are 12 or more terabytes of tweets everyday on Twitter (Easton, 2012) in which theatres can use hash tags to search mass amounts of data of different reviews from critics or audiences on variety of shows and performances.

B. Variety

Structured and unstructured data come in a variety of formats such as social, documents, images/videos and even voice recordings from computers, tablets and smart phone devices (Rijmenam, 2013). Large organisations and even SMEs have a variety of data sources that could bring opportunities in improving the efficiency and businesses (Dumbill, 2013). According to Gardner (2014a) theatres spend a lot of time
collecting data on audiences. Theatres normally monitor social media, read critic reviews and check box offices sales as a variety of data available to improve the quality of their performances (Smith, 2010).

In 2010, UK theatre industry bodies such as The Society of London Theatre, Theatrical Management Association and Independent Theatre Council came up with a new format of collecting data which is questionnaires (Smith, 2010). The new formats of collecting data were handed to audiences at the end of a performance to measure their emotional response (Smith, 2010). This can help theatre companies know if performances are a success or a failure and improve diverse performances linking with different age groups and interests.

Mostly major theatres have membership schemes which theatregoers pay yearly giving benefits for tickets and other promotions to the member (Gavin, 2008). These membership schemes are what many theatres businesses dependant on (Gardner, 2014b). This gives an advantage in order to use a different format of collecting data knowing different audience’s interest in providing special offers to increase more people to view performances. It also gives a chance for theatres to develop and expand the relationship with the audience (Gardner, 2014b).

Other sources of big data are also collected in reward systems such as loyalty cards from different industries. Supermarkets like Sainsbury’s collect large quantity of data from customer’s nectar cards and produces an outline of the customer on what they buy and spend (Ferguson, 2013). With a Nectar card customers gather points with every purchase and receive offers for shopping, dining or use the collect points for next grocery shopping purchase to save money (Nectar, 2002a). In 2011, Ambassador Theatre Group (Britain's largest theatre operator) developed the first loyalty reward card scheme for theatre audiences (Trueman, 2011). Theatre can use this as an opportunity like supermarkets to gain the audience's profile on which shows they attend and what kind of performances they watch.

Theatres can have a big opportunity of using all of these variety sources and combined into one by developing a mobile app for audience’s smart phone. For example, Sainsbury’s Nectar card has its own mobile app for clients to have easier access to their account and check their nectar points/offers (Nectar, 2002b) According to Kelly (2013) mobile devices is the future for analytics and can provide a new tool creating big data and give insights. By incorporating a mobile app, theatres can combine theatre membership account for audiences in order to gain easy access to details. Moreover, audiences can view special offers for tickets and answer questionnaire/survey reviews for different performances which they have seen around the UK to give feedback for improvements.

C. Velocity

In this digital age, the speed of internet is increasing rapidly and data is generated continuously at very high speed (Normandeeau, 2013). Since social media and box office reviews data is being used by theatre the growth and velocity of data increase as well giving an opportunity for real time data. Twitter generates 100,000 tweets with 700,000 Facebook posts every minute and any of these high speed data/information can be related to audiences reviewing theatre performances (StateTech, 2013). For instance, 40% of tweets from Twitter are linked with television which is being applied in TV ratings (Grimes, 2013).

From real time high speed data according to Ruggenberg (2011) studies show that 35% of positive reviews from social media give audiences a reason to go to theatres. The velocity of social media data plays a central part for audiences to find information about the theatres especially with the primary source of word of mouth (Mermiri, 2013). In particular, with one in five people writing reviews of performances/shows they have seen (Mermiri, 2013).

D. Validity

According to Lilley & Moore (2013) the arts and cultural sector's approach in the use of data is outdated and insufficient, failing to make most of the value of the data. As mentioned theatres used the questionnaire method to measure and gain knowledge of audience's emotional response of the performance they attend can bring valid/accurate data from the audience itself (Smith, 2010). In addition, theatre use loyalty card system to generate and illustrate the value of data of the audiences, theatres can use this data/information creatively and gain better use of the data (Bowden, 2014).

These data responses from the questionnaire and reward systems gives an opportunity for a new way into bringing validity of collecting data besides using current data collection methods such as social media, box office etc. (Smith, 2010). This can help theatres to make the right decisions based on evidence to improve the success of different performances (new or current) and help improve a variety of specific target marketing.

E. Value

Theatres have many opportunities and benefits with big data especially gaining large volumes of data from a variety of sources as mentioned, however, if there is no value in the data then it's useless to theatres (Marr, 2014). There is an estimate of one-third of theatregoers during a performance use their mobile phones while one in five theatregoers use social media to write a review on performances they have seen (Smith, 2013). This brings an opportunity in capturing the value of data as UK theatre industry bodies came up with a system in form of questionnaire and hand them to audiences at the end of a show, audiences will much prefer to carry the questionnaire and hand them to audiences at the end of a show, audiences will much prefer to carry the questionnaire through social media or through a mobile app. Especially since there are new studies from TicketMaster notice a increase of younger 16-19 theatregoers (Smith 2013 cited in Mermiri 2012). Social media is a big influence and essential part of people lives (in particular with young people) (Digital Me, n.d.) this can gain more responses from the questionnaire. Furthermore, mobile apps have also been a big influence as part of people's daily lives, the overall mobile app usage has increase by 115% in 2013 (Khalaf, 2014). As mobile devices are exceeding personal computers (IBM, n.d.) it brings the value of data more personal from the audience which can bring a benefit and an opportunity to theatres (Ballve, 2013).
III. THE CHALLENGES/RISKS OF USING BIG DATA FOR THEATRES

Big data has its opportunities which are mentioned looking through the characteristics of big data Vs to improve theatre in bringing more audiences to see performances and using theatre's creativity when analysing big data to produce value into the business (McDonnell, 2011). However, there are also challenges within exploring the big data Vs which will be explored in this section, and theatre needs to consider these challenges if using big data to overcome these issues.

A. Volume

Receiving large amounts of data can bring opportunities for SMEs especially in theatres; however, there are a few challenges when dealing with mass volume of data. Especially in social media sites, 90% of data are basically unstructured (PC Quest, 2013), however, to structure the data can cause an issue for the theatres as there is a lack of people with little analytic skills (Lilley & Moore, 2013). With increasing amounts of data, cost is a main challenge when storing unstructured data (Inbar, 2013). Having large amount of data doesn't necessarily mean that the theatre needs to store it all. However, there are theatre organisations which can provide funding such as R&D Fund for the Arts who can support big data in which won't have an issue of cost (Digital R&D, 2012).

Although it has the opportunity of increasing jobs to analyse and organise large volumes of data, yet, there is a challenge of SMEs especially in theatres having the lack of skills with big data (MacInnes, 2013). 57% of SMEs sectors are showing concerns which are suffering complications when employing experts in big data (Passingham, 2013).

B. Variety

Businesses points out that the type of data being captured has also changed, from “structured data”, which can be easily analysed, to “unstructured data” such as Facebook updates and YouTube videos which are harder for computers to decipher. (Taylor, 2012a). It’s easier to have the data in the same format however; the challenge for theatres and other organisations in using different data formats is not the case. Many different types of data has different time periods in analysing, for instance, as theatres collects sentiment emotions from theatregoers in social media (Rijmenam, 2013). 85% of businesses are producing more unstructured data which can cause difficulties to the producers and directors of the performance when deciding on improvements in trying to make sense out of the valuable data and time consuming (Plummer, 2014).

Another challenge that can cause issues to theatres is privacy especially when data is collected from loyalty cards or a mobile app and theatregoers aren’t given permission. This can provide a conflict with UK Acts in particularly the Data Protection Act (1998) and Human Rights Act (1998). To ensure this challenge doesn't arise theatres can apply guidelines as big data is an emerging topic and let theatregoers know how their data is being used and consider legal laws in order to keep their audiences safe (Goodendorf, 2013).

C. Velocity

A challenge which can cause a risk to theatres is having unstructured data from social media which are difficult to analyse means the speed of analysing the data can lead to being constricted (Taylor, 2012a).

D. Validity

The accuracy of data can be challenge especially when it comes to audiences' interests (as their interests may vary), if theatres are collecting data using valid questions having more data doesn't necessarily mean the data is accurate (Taylor, 2012b). Even though theatres are using questionnaires systems to gain improvement on other sources of data, does the questionnaire provide ultimate answers to help develop theatre performances? That answer is uncertain and don't know if it's valid (Smith, 2010). According to Gardner (2012) explained overall theatres are rather slow when it comes to huge changes in social media. This gives another challenge as if theatres are not updated to changes in social media, how does the theatre know what they read is valid (Gardner, 2013). The amount of available data that exists today and the rate in which is updated leaves the validation of data too complex for us to understand which is a challenge to SME theatres (Nessi, 2012).

E. Value

This can provide a challenge to theatres when analysing which data is valuable to increase audiences and improve performances. This is crucial part of big data to any businesses extracting the value from the data, if there is no value then the data is useless and will lead to wasting funds especially theatres using funds from third parties such as Digital R&D Fund for Arts (Tallon, 2013). Another challenge to theatres value of data is social media, if audiences and theatregoers use social media when reviewing a performance. According to Kardamaki (2013) brands from large enterprises ignores 70% of the comments that customers publish which is a waste of valued data. Theatres need to consider that big data has a harmful implication in the arts and cultural sector as where audiences' emotions and personal choices are highly valued (Brown, 2013).

F. Verification

Theatre-goer's interests vary when it comes to performances which lead to a growing uncertainty of data. As theatres are using social media 80% of the available data will be uncertain, so can theatres trust the data available? While providing questionnaires through social media as an opportunity there is a risk of people lying when completing a questionnaires/surveys. An example of this is a female lying about her identity for two and a half years dressing as a male (Milligan, 2012).

A way to overcome this challenge is to have questionnaires/surveys to theatregoers with loyalty card membership and social media groups in order to complete the questionnaire they can have the chance to be automatically entered in a competition and a person will be selected randomly to win free tickets for a performance as an example. This can help theatres to collect specific trusted data on audiences and reviews from performances they have seen so far.
which is also valuable. Another challenge to theatres as mentioned before is that they are vague when it comes to new changes in social media, this brings an disadvantage to theatres if they are not up to-date with current social media sites, leads into an interesting question of what audiences wrote if it can be trusted from social media (Gardner, 2013).

IV. CONCLUSION

Big data has many opportunities and benefits to SMEs especially in theatre. Providing opportunities to increase the number of theatregoers enhance performances and improve target marketing for specific audiences from collecting variety sources of data and analysing the data to bring the value. However, before using big data into theatre need to address the challenges and consider before using the emerging big data technology. SMEs theatres should use big data to help increase audience numbers, improving new performances based on relevant data and leads to being ahead of other theatres by giving it a competitive advantage.

V. REFERENCES


Analytics in a Mobile App for: 30

To Describe


Are You Really Ready For Big Data?
An Explanation Of The Skill Shortages For Big Data In SMEs

Wesley Simms
Business computing and Law
University of Derby
Derby
UK
W.Simms1@unimail.derby.ac.uk

Abstract—in the last few years, Big Data Analytics and strategies have been adopted by many large Organizations. Now smaller to medium sized enterprises (SMEs) are getting on board. However research predicts that there will be a large skill shortage within the field of big data causing some companies to lose business. This paper identifies the main skills that executives are looking to bring to their departments, pointing out some job titles/positions that will be widely known in 2018.

Keywords—Big Data, Data Scientist, Analytics, Data-driven, SMEs, Skill shortage.

I. INTRODUCTION

Traditionally, data would be analysed manually and interpreted in order to retrieve valuable information. However with the huge amounts of data being analysed by modern organisations, this approach becomes too time consuming and expensive (Fayyad, Piatetsky-Shapiro & Smyth 1996).

Big Data cannot be defined in a specific size (megabytes, terabytes etc.), it is predicted that as technology grows, the size of data stored will also continue to grow (Manyika et al. 2011).

Two of the largest difficulties in big data analytics are finding and creating useful combinations of relevant data to meet company needs, as well as managing an unimaginable amount of this data (Bizer, Boncz, Brodie & Erling 2011).

The rapid shift across many organisations and the impact on their technological needs creates concern for executives in preparing their companies for dealing with big data analytics (Bughin, Chui & Manyika 2010).

II. WHAT IS BIG DATA?

The key feature that makes big data so large is the use of repeated observation over time and space (Jacobs 2010). Many companies are still unsure about the costs and benefits in big data projects (Gopalkrishnan, Steier, Lewis & Guszcza 2012).

In fact, some studies identified that many SMEs have little or no knowledge about big data analytics at all (E-SkillsUK 2013a). Big data will help small businesses by enabling them to use the insights and capabilities of large data sets which only certain large organizations may have had access to in the past (INTUIT 2012). Retail companies compete on a basis of analyzing consumer needs and trends, then making suggestions and recommendations based on similar products. Also, attempting to provide cross industry price data in order to display discounts to customers effectively is another form of analysis that companies use (Bughin, Livingston & Marwaha 2011). The scale of these projects requires both a lot of product observation, over geographic location and possibly over long periods of time. This type of workload would benefit from the use of big data analytics (Jacobs 2010).

III. THE SKILL SHORTAGE

Too few executives understand the business potential and assets that big data brings to the table (Bughin, Livingston & Marwaha 2011). Similarly the lack of staff and skills in big data analytics is known to be one of the biggest barriers for companies looking to get involved in the field (Russom 2011). The lack of skills and understanding of big data can lead to non-uniform distribution of data. Non-uniform distribution is a common and large issue in big data analytics. This can seriously impact the time taken to process the data in question (Jacobs 2009). Therefore the implementation of successful and equal distribution of data across networks is an important skill to possess. A great deal of organizational data is largely unrecognized, underutilized and inaccessible. Data should ideally be considered a corporate asset to organisations (Brancheau, Janz & Wetherbe 1995). However this is not yet the case for many organisations. The implementation and use of information based architectures are a highly ranked issue among organisations (Brancheau, Janz and Wetherbe 1995).

The implementation of such architecture however requires the appropriate skills and understanding. The infrastructure of a system or network will not be responsive if data is scattered around the system without a viable plan. This affects software’s ability to integrate and distribute itself across networks. Successful information architectures would solve these issues (Brancheau, Janz & Wetherbe 1995). McAfee & Brynjolfsson (2012) have identified five management challenges to becoming effectively data-driven. The challenges are in leadership, talent management, technology, decision making and company culture. Without effective management in these areas businesses may not fully benefit from the implementation of big data analytics.

IV. WHAT DO ORGANISATIONS WANT?

Organizations are looking for new initiatives that can determine what technologies can be used at a reduced cost and
provide greater usage to their customers/clients (Neckopulos & Hansen 2012). Organizations are also trying to find new opportunities for revenue and growth, meanwhile improving risk management (Neckopulos & Hansen 2012).

Among the Top 10 issues of organisations in the United States, the development and implementation of an Information Architecture came in fourth place (brancheau, Janz & Wetherbe 1995). The effective use of data as a resource came up in 7th place in the same survey (brancheau, Janz & Wetherbe 1995).

McAfee & Brynjolfsson (2012) believe that too many executives currently rely on intuition and experience, and less on their data. A study that they conducted in which they asked a number of executives how they rated their companies at being data-driven on a 1-5 scale. 32% of responses rated themselves at 3 or below (McAfee & Brynjolfsson 2012).

Many executives report customer relationships and information about sales to be large priorities. However they believe that faster access to data will help them keep up with the service expectations held by their customers (Avanade inc 2010). This simply puts more data in their hands in a shorter space of time. This complicates the problem as 56% of business and IT executives feel overwhelmed by the amount of data their company is responsible for. In addition, 1 in 3 executives cannot find the right people who can provide the relevant information needed at the right time (Avanade inc. 2010). This resulted in more than a quarter of executives losing business due to being unable to deliver the correct information where necessary (Avanade inc. 2010). In short, organisations are concerned about their data and what it means to them, however many feel they lack the skills within their organisation to deal with, or properly interpret the data. This leaves work to experience and intuition, which may harm the growth these companies wish to pursue.

V. CONSEQUENCE OF THE SKILLS SHORTAGE

United States will face a need of 190,000 new analytically skilled workers by 2018. In addition 1.5million managers and analysts who can interpret and make effective use of big data will be also be needed (Bughin, Livingston & Marwaha 2011). Companies with revenue of $1billion and over a year believe that they are losing around 13% of this due to poor data management (Oracle 2012). This translates to over $130million a year! The use of big data analytics will cause change across organizations, in certain areas of industry, more significant changes may be needed. The lack of skills in the area of Analytics will make these changes difficult. Without appropriate change in organizations, newer and possibly smaller competitor companies may gain leverage in the market (Manyika et al. 2011), 46% of companies revealed that they have made inaccurate business decision due to misinformed or outdated data (Avanade inc. 2010). Unless companies invest in the knowledge and understanding of big data analytics, it is likely that these companies will continue to suffer difficulties of this magnitude and continue with financial losses.

What this means for SMEs

Innovation within SMEs can typically be hampered by a lack of resources and accessibility to specialised workers. This can make it difficult to advance or gain business in their sector (van de Vrande, de Jong, Vanhaverbeke & de Rochemont 2009). Large organisations over the past decade have been working with many SMEs in the outsourcing of projects due to the flexibility that SMEs provide (Narula 2004). This makes the SME’s job of networking with larger companies a very important task as this allows SMEs to overcome the disadvantages of their size and adopt a philosophy of open innovation (van de Vrande, de Jong, Vanhaverbeke & de Rochemont 2009).

The shortage of skills in big data may require some of the larger companies to outsource their data analysis needs. With many other SMEs having little to no knowledge of big data (E-SkillsUK 2013a), it may be wise to develop these skills in order to meet the demand for skills in big data analytics (Bughin, Livingston & Marwaha 2011). With these skills in place it would be possible for SMEs to relieve the pressure from their larger counterparts while maintaining secure working relationships themselves.

VI. WHAT SKILLS ARE NEEDED TO UTILIZE BIG DATA?

Companies need to find scientists and individuals with the ability to discover patterns within data and develop appropriate business information from these findings (McAfee & Brynjolfsson 2012). The specific skills that are going to be important in big data Analytics are as follows:

- Leadership - The need for clear goals and human insight will not be erased by big data. In fact, effective leadership in this time of organisational change will be a decisive trait in many successful companies (McAfee & Brynjolfsson 2012).
- Analytics – with a projected Talent gap of 50-60% in deep analytics skills, the ability to analyse and make sense of large amounts of data will be in high demand by 2018 (Manyika et.al 2011).
- Communication – Effectively communicating the findings from the data will also be an important skill for organizations to look for.
- Programming – The need for development experience in languages such as NoSQL, Java, Javascript and SQL and MySQL has risen by over 673% per year over the last five years (E-SkillsUK 2013b).
- The proper utilization of these skills will equip organisations with the knowledge and skilled workers needed to undertake big data projects. However simply having the skills may not be enough. For a company to succeed they may need to adopt a new philosophy of data, or a data culture.

VII. COMPANY CULTURE/ DATA CULTURE

A company culture which focuses on data requires a philosophy across more than one department. The purpose of this new way of working is to enforce in companies the importance that their data carries. Previous business models and way of working do not offer the correct features for big
data analytics to take place (McAfee & Brynjolfsson 2012). McAfee and Brynjolfsson (2012) identify a number of questions that executives should start asking when dealing with company data: “What do the data say?”, “Where did the data come from?”, “How was the data analysed?” and “How confident are we in the results?” Questions like this allow executives and analysts to have their intuition and experience overruled by the results of this data. Organisational change and the adaptation of a “Data Culture” is mentioned across multiple research papers (McAfee & Brynjolfsson, 2012; Avanade inc 2010; Manyika et al. 2011, pg. 12). A Data Culture suggests that each worker, be they an analyst or executive be mindful and knowledgeable about data and its importance to business innovation. For this to work effectively it would require a greater depth of knowledge in analytics and statistics across departments. At the very minimum it would require an attitude that suggests the value of data being the top priority across these departments and the creation of a competitive company structures (Manyika et al. 2011).

VIII. DATA SCIENTISTS

Data scientists, this will be job role that will be heard about a lot more in the near future (Davenport & Patil 2012). Data scientists will be people who specialize in making discoveries while analysing data, bringing structure to previously formless bundles of data (Davenport & Patil 2012). Skills that they will possess will include: programming, analysis and most importantly, the ability to communicate verbally, visually and in writing (Davenport & Patil 2012). However Data scientists will only fill a short proportion of big data related jobs in the industry (E-SkillsUK 2013a). Some other job titles to look out for that may involve skills in big data Analytics include: Business Analysts, Data Analysts, Engineer, Research Analyst, Research and Development Specialist, Business Intelligence Directors, Business Intelligence Specialists (Russom 2011). However useful analytic skills are not solely based upon these job roles. Some other job titles that also use analytics would be Consultants, Marketers, Risk Managers, Statisticians and Data Governors (Russom 2011). These examples may help provide an idea of what beneficial experience future employees will ideally have.

IX. BUSINESS INTELLIGENCE TOOLS

Current Business Intelligence (BI) tools and software typically do not have the statistical functionality necessary for effective analysis. A look into new software packages that allow for statistical analysis and provide massively parallel databases may also be beneficial to companies (Cohen, Dolan, Dunlap, Hellerstein & Welton 2009). The causation models and scientific models formerly used to manage your data may no longer be sufficient. It is suggested that the deduction of data from applying theory can identify and develop useful correlations (Bollier 2010). This allows data-analysts to make predictions about possible correlations and allows them to learn from the data itself, instead of hoping to find specific results on a scientific model (Bollier 2010). Skill in this form of deduction and analysis will be very useful to companies in the next few years. This shows that the software and business intelligence tools can be help companies in handling big data. However software alone will not necessarily allow organisations to fully incorporate big data analytics; for this software to produce meaningful results a certain amount of human intervention is required.

X. BRIDGING THE SKILLS GAP

In order for companies to be successful through the use of big data it is necessary that they focus resources into developing and maintaining a good knowledge of their data and the skills required to use it to its full potential (Avanade inc 2010). Businesses must maintain clear and well defined business goals; meanwhile constantly assess these goals in order for their analytical processes to encompass big data (Gopalkrishnan, Steier, Lewis & Guszcza 2012). Skills in leadership, programming, analytics and communication should be prioritised in training and employment to compensate for the predicted shortage (McAfee & Brynjolfsson 2012). One of the more important challenges that SMEs may need to overcome is the philosophical change to focus on the meaningfulness of the data they wish to examine. Statistical, business and economic considerations should be made accordingly to how meaningful the data are and how effective it will become to the organisation after analysis (Gopalkrishnan, Steier, Lewis & Guszcza 2012). It is possible that new software and database solutions could help companies adapt to the change (Cohen, Dolan, Dunlap, Hellerstein & Welton 2009), however it is probable that the best way to truly prepare for big data is through theoretical deductions and statistics as opposed to feeding data through existing evaluation models (Bollier 2010).

To summarise, the development of knowledge in big data analytics, use of clear and concise business goals, appropriately trained and skilled staff and an alteration in company philosophy will allow companies to successfully conduct business in big data analytics.

XI. RELATED WORK

Many companies believe that the use of big data will bring them a better understanding of what their customers want out of their organisation. However for certain businesses, this may not be the case. Therefore further research into customer relationships management (CRM) could provide a simpler solution until big data analytics becomes a conscious goal for the organisation (Chen & Popovich 2003).

For companies interesting in learning more about big data, research into data mining would be beneficial as this subject targets the approach and even techniques for gathering and storing vast quantity of data (Berry & Linoff 1997).

As mentioned earlier, software and the use of business models are not the only approach to take when tackling big data analytics. The skills and techniques required are largely human. Research into augmented intelligence could provide a new scope on the ways data can be analysed. Schmitt (1998) identified that the use of augmented intelligence can be used effectively in designing physical architectures. In addition the combination of business intelligence and competitive
intelligence may also be of interest. Competitive intelligence focusing on the external competitive environment the organisation is part of, in particular the analysis of data from rival/competitive companies (Zheng, Fader & Padmanabhan 2012).

XII. CONCLUSIONS

It is evident that, in the next few years, many organisations will be undergoing extensive changes in order to support this new business trend that is big data analytics. As identified, the predicted skill shortage is not an issue that is going away overnight; however it is one that will be of considerable challenge. Some businesses are going to have to make large internal changes in order to keep up with the new trend that is big data. It will be necessary to adopt new training methods, develop focus on skillsets in analysis and to create a company culture around the importance of their data. Success in big data analytics may present smaller and medium organisations with opportunities for growth, but a resource that will be of considerable advantage during and possibly after the skill shortage, while others may continue to experience losses in data and in business.

XIII. REFERENCES


Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) From Data Mining to Knowledge Discovery in Databases. AI Magazine. 17(3).


Security Concerns Are Deterring SMEs From The Adoption Of Big Data Analytics

Matthew Smith
University of Derby
Derby, United Kingdom
m.smith24@unimail.derby.ac.uk

Abstract— As large scale companies increase their reliance on Big Data analytics, SMEs are becoming increasingly attracted to Big Data and the financial advantages it brings. However, the fear of a data breach and the repercussions that would follow are deterring these companies. This report investigates a selection of areas that are causing SMEs to abandon Big Data, looking at the reasons behind the fears and what possible solutions are available. As possible solutions are investigated, these areas could reduce the fear associated with Big Data by SMEs, allowing for the increase in use of Big Data by these companies and the financial gains that follow.

Key Words— SME, Big, Data, Analytics, Breach, Target, Compliance

I. INTRODUCTION

As big data burst upon the scene in the first decade of the 21st century (Davenport, 2013), the only organisations brave enough to embrace it were large online companies and start-up firms, using the data collected to improve their businesses and increase their fortunes.

This ‘Big Data’ - a collection of data from traditional and digital sources inside and outside of a company that represents a source for ongoing discovery and analysis (Arthur, 2013) - is an essential tool for companies, allowing them to utilise the wealth of information collected and created by monitoring the activity of users and sales, allowing this data to then be fed back into the company, allowing for improvement in areas to be made based on hard analytical data.

Companies such as Google, eBay and Facebook all rose in success due to their use of Big Data analytics, mainly due to Big Data playing such a large part in the infrastructure of the companies. As Davenport (2013) mentions: “[the companies] didn’t have to merge big data technologies with traditional IT infrastructures because those infrastructures didn’t exist. Big data could stand alone, big data analytics could be the only focus of analytics, and big data technology architectures could be the only architecture.”

As these online companies and start-ups began increasing in value, other industries began seeing the importance of having Big Data analytics as a tool to support their businesses. Companies were attracted to Big Data due to the opportunities it presented and also the low cost of the technologies involved (Davenport, 2013). The large scale organisations, with their substantial budgets allocated towards security and privacy, were able to implement Big Data solutions that also ensured the relative security and privacy of the data they were collecting.

In order to keep up with the competition, small and medium enterprises (SMEs) were forced to adopt Big Data analytics into their businesses, using the gains of Big Data to help bridge the gap between them and the competition. It has been estimated that if SMEs matched the growth rates of the market leaders, SME revenue could grow by $770 billion just in 5 select countries, with an estimated 6.2 million extra jobs being created across the globe (Michael et al, 2013). However, with this adoption of Big Data analytics comes the realisation that they too have to meet the same levels of information security compliance as the larger organisations – just without the budgets to match. This has resulted in data security being labelled the ‘SME’s top technology concern’, according to a report by the Boston Consulting Group (Joshi, 2013). With data security covering such a vast area, the rest of this report shall investigate these areas of concern as well as provide recommendations on how to solve them, allowing SMEs to adopt Big Data analytics without the fear of data breaches or attacks or the implications that come with them.

II. ISSUES SURROUNDING SMEs

A. Lack of Knowledge and Training

One of the main fears within SMEs that is preventing the acceptance of Big Data analytics is the risk of a security breach, and especially the implications that follow. However, one of the main causes of data breaches is a distinct lack of knowledge in the area by the staff themselves. This lack of knowledge needs to be solved before the implementation of any Big Data analytics. In a whitepaper released by WatchGuard, the company reveals the top ten threats to data security within SMEs. These findings reveal that the main security risks to SMEs are areas that human error plays a key role, although other security risks are also an issue.

Due to the smaller budgets of SMEs, the report found that corners were often cut to save time and money, resulting in weaknesses within the system, allowing for attack. The use of smaller, local companies or even amateurs to install networking equipment was given as an example, with the lack of experience resulting in basic security issues being solved such as router configurations being changed from default (Pinzon, 2008). In a report by Verizon, this finding was supported, with 62% of 500 real-world breaches survey being the result of
internal errors, such as the changing of router configurations (Baker et al, 2008). The development of websites using custom, poorly written code was also found to be a major target to hackers. A low paid, low experience member of staff without knowledge of these security issues and their accompanying solutions could write server-side code that allowed for the use of SQL-injection to compromise the websites, both internally and externally (Pinzon, 2008). Major companies that have experienced this type of attack include Sony (Danchev, 2008) and the British Government (Athow, 2008). These websites are then used to spread malware across the computers that access it, allowing backdoor entry into their systems through these infected PCs.

An overall lack of training and experience within employees was discovered to be the result of many of the ten threats to companies, with employees being unaware of the security risks that may allow for an attack on their company’s system, and the subsequent loss of data. The use of unsecure hotel and public Wi-Fi access points was found to be one of the major concerns. These networks were found to be often insecure and traffic was not encrypted, allowing for the use of packet sniffing tools to access usernames and passwords from devices using it (Web Admin Blog, 2008). The report continues with the fact that this type of attack leaves no trace of compromise on the victim’s computer, so is very hard to detect without prior knowledge (Pinzon, 2008). The loss of un-encrypted devices and the “reckless use of the web” were also found to be key areas of weakness for companies, both of which are preventable through education of staff members, yet still account for a major part of the wider threat of a data compromise.

B. Security Fears of Using the Cloud

Despite the wide range of benefits of using the cloud to store and process Big Data, including flexibility and reduced hardware investments, SMEs are still weary of using it, mainly due to security concerns. It was recently revealed that 56% of senior-level security and IT respondents admitted that security concerns have led them to no starting or finishing cloud or big data projects (Dark Reading, 2013). With large enterprises, often with over 5000 employees and dedicated technology staff abandoning the cloud and big data due to security, SMEs have an even greater challenge utilising Big Data and the cloud.

One of the fears is putting all of the data in the hands of others, resulting in a greater risk of data breaches. According to a report by Computer World (2013), any data uploaded that is encrypted and stored is encrypted using keys that are held by the cloud staff, not the owner of the data. This results in a “rogue” employee having the ability to decrypt data and leak it. This is one of the main fears of SMEs, trusting a third party to store the data to save money and reduce the need for dedicated hardware, yet increase the risk of data being leaked by the cloud storage company in question.

According to a report by Computer Weekly (Kelly, 2011), there has been no major security breaches on the cloud so far, although “it could only be a matter of time as cyber criminals wait for the right moment to strike”. This is supported by material stating that DDoS attacks against cloud providers have been able to disable their firewalls and prevent access to the service through the use of Botnets, all sending traffic to the cloud servers to overload their connections (Cobb, 2013). For SMEs that cannot afford dedicated hardware and resources to deal with their Big Data, the cloud is a cost effective alternative, yet it brings increased risks to both security and privacy of any data stored upon it.

C. Compliance

Another area that overwhelms SMEs is the level of compliance that they have to reach whilst handling and processing Big Data. Government sets levels of compliance are enforced for any companies handling data that is of a personal nature, such as the type of data regularly seen as ‘Big Data’. These levels of compliance vary between countries and regions, but it was found that of 100 SMEs surveyed, 22% were not compliant and a further 14% did not know if they were compliant or not (ITWeb, 2014). Those that do not meet levels of compliance are subject to financial penalties and litigation if data is compromised as a result of this (ITWeb, 2014).

However, small and medium companies have an ever increasing challenging meeting these levels of compliance necessary, such as the Data Protection Act and PCI-DSS (Kelly, 2011). These overly complex measures, which include completing a 48 page document with “the best part of 400 questions” (Kelly, 2011) have resulted in a large amount of SMEs being put off the use of Big Data, with meeting levels of compliance an unnecessary hassle. Ensuring that compliance is met allows for standards like ISO 27000 (ISO, 2009), to also be met and complying with these ensures that customers and users entrust the service that is provided.

D. Financial Impact

One of the biggest fears to SMEs is the negative financial impact that they would receive if a data breach occurred. According to PwC’s Information Security Breaches Survey (2013), the average cost to even a small company for its worst security breach of the year is within the range of £35000 and £65000. This rises significantly to between £450k and £850k for large companies (PwC, 2013). The report then goes on to state that 87% of small businesses had a security breach within the year, up from 76% in the previous year. With the rate of attacks increasing every year, ensuring that protection against these attacks is essential. With the median amount of attacks per year being 17 for a small company, the financial impact of this would destabilise any company (PwC, 2013).

However, it is not just the direct cost of data breaches that affects companies. According to a data breach report conducted by Ponemon Institute (2013), there is a 23% chance of loss of customers and business partners following a single data breach. There is also a 25% chance of a change in public opinion, with negative reports by the company and also members of the public. Both of these, followed by a 16% chance of regulatory fines, all would result in a negative impact on the financial status of the company, from loss of sales to fines and restrictions by the government (Ponemon Institute, 2013).
These direct and indirect financial impacts are frightening to companies, both large and small and with the risk of data breaches at the highest level to date (PwC, 2013), SMEs must ensure that data breaches do not take place, as just a single data breach may have devastating financial consequences.

III. CASE STUDIES

A. Target

A recent, large scale data breach that serves as an example for all the points raised so far is the recent hacking of US Chain Target, in which millions of credit card details were stolen. Over 100 million credit card details were stolen using malware placed within Target’s systems. This all came from a company that was PCI certified and had a security system worth $1.4 million (Business Week, 2014). Since the breach, 90 lawsuits have been filed against Target and the company has spent $60 million. Profits fell 46% as a result of customer distrust (Business Week, 2014). However, it was not just Target that felt the consequences. Banks and Credit Unions lost $200 million as a result of the data breach (Cross, 2014), due to the reimbursement of funds and also the replacement of credit card details stolen, at a cost of $10 each (Mead, 2014).

This is a perfect example of why SMEs are afraid of Big Data. With large scale companies, running multi-million dollar protection systems, being hacked and data being breached on a massive scale using fairly simple techniques, it speaks a message to small and medium enterprises that even the large companies are not invincible.

B. Specs Liquor Chain

Another example of a case where financial data has been stolen from compromised systems is Specs, a US liquor chain. Much like Target, Spec’s’ system was compromised for a lengthy period of time, with the initial breach taking place on October 31st, 2012, and has continued as late as March 20th (Kaplan, 2014). During this period of time, more than half a million customer’s financial data was been stolen from 34 stores. During this period of time, malware was installed on the Spec’s computer systems, recording the financial details of customers paying by card.

Experts brought in by Spec’s admit that the security measures in place were “clearly ineffective or inadequate” (Kaplan, 2014), highlighting the problem that even major chains experience data breaches due to poor security management and protection. The inability to spot the breach within the 17 months it took place highlights the hole in security systems that were previously thought to be secure, as new and more sophisticated methods of breaching systems are created, targeting systems that are still running these outdated security systems.

IV. POSSIBLE SOLUTIONS

With SMEs fearing Big Data due to mainly a lack of knowledge and expertise, the solution to this lies in the educating of staff and company leaders of the benefits, but also the risks of Big Data and what is to be expected if the implementation of it goes ahead. The report by BCG (Michael et al, 2013) states that companies should hire senior executives skilled in IT rather than relying on local IT vendors, who are less likely to be adequately skilled. This would solve the issue of education, with having one highly skilled staff member managing Big Data and ensuring compliance is met. This member of staff could then recommend and educate the others, on matters such as the use of public Wi-Fi spots as previous mentioned.

However, the likelihood that these SMEs can afford the cost of hiring a dedicated member of staff for security and data management is unlikely, especially in companies with less than 50 members of staff. It is for this reason that security and data compliance is normally managed by a senior member of staff at the company already.

Many companies are still unaware of the problems they face when implementing Big Data within their companies, mainly due to the lack of education that they receive from both the government and from other companies and organisations. This has been highlighted in the previous sections of this paper, but solutions available to remedy this situation are small in number. Symantec recently found that 31% of all targeted data attacks were aimed at SMEs with less than 250 employees (MacInnes, 2013). It has also been found that SMEs are targeted in order to target larger, connected companies as part of a larger goal. This is similarly reflected in the case of Target, where a small-scale HVAC company was targeted in order to breach the Target computer system. ‘Watering-hole’ attacks such as this are becoming more common according to the latest reports (Feinberg, 2014).

One of the most recommended solutions to this issue for SMEs is to use ‘managed security services’, outsourcing the security to companies that are aware of the latest security flaws and attacks taking place. As SMEs have wildly varying needs, this solution can accommodate to this with a combination of on-site and managed services, depending on what is required by the company in question. This is also a more affordable solution, with no need to pay staff when not needed. With complicated and demanding requirements for data compliance being a major issue, a solution to this is also to adjust the policies that require data compliance so that they are more suited to SMEs. Applying the same policies to both large and small companies pushes away the smaller companies. The report by BCG mentions this, calling for certification of vendors so that a minimum level of security, privacy and data protection are in order (Michael et al, 2013). The changing of certification or creation of less-stringent certification for SMEs would allow for the uptake of Big Data without jumping through so many hurdles, a time consuming procedure according to Kelly (2011).

V. CONCLUSION

Despite the tempting advantages of Big Data to SMEs, there are many fears regarding the method of analytics that are holding back companies, leading to data security being labelled the SME’s top technological concern (Joshi, 2013). These concerns have all been detailed in the previous sections of this report, including the fear of attack and the possible
consequences of a possible data breach. The hurdles created by the government in order to store and use Big Data have also been revealed to be holding back companies, due to the complexity of the process of becoming certified.

Two case studies have been investigated as part of this, Target and Spec’s, which both have recorded damaging data breaches as a result of inadequate security protection, highlighting the fact that even major companies are vulnerable to attack when security measures are not enough.

Possible solutions to some of the problems have also been discussed, as mentioned by the BCG report on Big Data in SMEs, amongst other sources. These solutions call for a change in the process of certification, making it friendlier to small companies that lack dedicated members of staff to monitor information security. The training of staff was also discussed, informing them on best practices for data security, preventing the risk of data breaches by eliminating any form of human error, one of the areas found to be the greatest risk to security (Pinzon, 2008). If these solutions were applied, the fear of attack would be far less. With education, data breaches can easily be prevented, securing the data and reducing the likelihood of any negative impact upon the company as a result of a breach. All of these would reduce the fear of attack and increase in the use of Big Data by SMEs, bringing them in line with the larger companies already using Big Data to their advantage.

VI. REFERENCES


Defining Big Data Analytics And Its Potential Benefits For SMEs

Josh Trow
School of Computing and Mathematics
University of Derby
Derby, United Kingdom
100196729@unimail.derby.ac.uk

Abstract—Big data analytics is the process of gathering and examining large amounts of data in the hope of uncovering meaningful patterns or correlations. SME’s (Small to Medium Enterprises) are able to benefit from big data in many ways, such as increased insight into customer purchase habits through social media feeds. Another example is identifying prime locations for premises based on location demographics from smart devices. This paper aims to establish what big data analytics is and discuss several of the ways SMEs can benefit from big data analytics.

Index Terms—SME, Big Data, Benefits, Analytics.

II. WHAT IS BIG DATA ANALYTICS

A. Definition

Due to the novel nature of big data there are a number of attempts at definitions across published material. IDC attempts to define it technically as a new generation of technologies and architectures, designed to obtain value from large volumes of data in real-time (Gantz & Reinsel, 2011).

Whereas Dumbill examines the practical issues in his definition, stating that big data is a reaction to the rise of data that exceeds the processing capacity of conventional database systems, which needs new technology to manage it. (Dumbill, 2012).

Finally, Gartner defined big data as large volume/variety data that must be processed in new ways to enable effective decision making strategies; he also defined key factors in the form of the three “V’s” (Laney, 2011). This definition is generally considered to be the comprehensive definition of big data, with the “V’s” concept used widely in big data discussion (Laney, 2001).

B. The “V’s”

Gartner originally established the three V’s in Laney’s 2001 paper as Volume, Velocity and Variety (Laney, 2001). Since then additional V’s have been added in order to provide businesses with a complete picture of all important aspects of big data. The additional V’s added more recently are Veracity, Variability, Visualization and Value (van Rijmenam, 2013).

- Volume – The sheer size of the data used in big data analytics can be daunting to some SME’s, with required storage ranging from terabytes up to Exabytes. An effective storage solution must be devised to house this data for analysis (The Economist, 2013). An effective storage solution must be devised to house this data for analysis (The Economist, 2013). An effective storage solution must be devised to house this data for analysis (The Economist, 2013). An effective storage solution must be devised to house this data for analysis (The Economist, 2013). An effective storage solution must be devised to house this data for analysis (The Economist, 2013).

- Velocity – Data in big data analytics is collected and analyzed at high speed, with companies often dealing with terabytes of data per day. This requires efficient and effective processing systems to handle the data (The Economist, 2013).

- Variety – SMEs work with a diverse range of data types, from structured to unstructured data. This requires sophisticated data analytics tools to handle and analyze the data (The Economist, 2013).

- Veracity – The quality and accuracy of data are crucial in big data analytics. SMEs must ensure the data is accurate and reliable to make effective decisions (The Economist, 2013).

- Variability – The data that SMEs work with can be highly variable in nature, with data sources and types changing constantly (The Economist, 2013).

- Visualization – Visually presenting data can be crucial in big data analytics, helping SMEs to better understand and interpret the data (The Economist, 2013).

- Value – The ultimate goal of big data analytics is to provide value to SMEs, helping them to make better decisions and improve their operations (The Economist, 2013).

This tough economic environment has meant that it has become increasingly important for small businesses to know who their customers are and how to appeal to them. In order for small businesses to compete with businesses on a larger scale it is very important to establish customer loyalty. 60% of consumers are willing to pay more for a similar product or service from a smaller business than they would from a large, corporate business (Blackburn & Giudici, 2013).

Big data is the logical next step in gaining insight into who a business’s customers are and their purchase patterns. By analyzing and collating data from a number of sources to find patterns and correlations, businesses are able to find useful trends which can be capitalized on. Despite this, as little as 0.2% of SMEs have a big data scheme, with 22% of SME owners reporting ‘poor’ or ‘very poor’ knowledge of big data (Wilkinson, 2013). Before discussing how big data can be beneficial to SMEs it is important to first establish a clear understanding of what big data analytics consists of.
• Velocity – One of the key benefits of big data is that its data streams are available in real-time. This means that businesses are able to see trends developing as they happen and capitalize on them. SMEs need to ensure that data is collected and analyzed in a timely manner in order for them to take advantage of the real-time nature of the data and not miss vital opportunities that may expire before they can capitalize on them (Oracle, 2013).

• Variety – In order to make the most of big data, trends need to be identified and links drawn between data sets. In order to do this a variety of data sources must be collected and correlations identified. For example, social media data may give a business insight into certain products’ popularity, but without relevant geolocational data it will be difficult to know where to advertise that product in order to reach the untapped market (Oracle, 2013). SME’s need a variety of data in order to build a complete picture of their customer’s habits and not risk incomplete data resulting in false trends.

• Veracity – In order for big data analytics to be effective for the business, it is of utmost importance that both the data and the analyses obtained from it are correct. It is important for SMEs to verify their data, as incorrect data can be catastrophic, with wasted resources being used for a market which does not exist (Maier, 2013).

• Variability – Data always needs to be examined in context. The meaning of a piece of data may be different one day than it had been the previous. For example if data was obtained via results of a search on twitter for a particular word (particularly hashtags), that word could change meaning overnight, going from positive to negative connotations, meaning that SMEs must keep up to date with current events to determine whether data is still applicable and ensure resources are not wasted (Katal, et al., 2013).

• Visualization – Due to the vast nature of the data sets involved in big data analytics it is often difficult to convey findings in sheer figures. Charts and graphs can prove crucial in identifying trends and being able to effectively communicate them across the business, or externally. Businesses need to be able to produce visualizations in order to view trends more effectively. Thankfully several third party services such as Ayasdi exist, which are able to produce these visualizations for a business, provided with the data (Rijmenam, 2013) (Ayasdi, 2014).

• Value – Finally, it is very important for businesses to understand the value of the data they gain through big data analytics. While the raw data itself holds no value, the analyses performed on the data can be extremely lucrative. For example the McKinsey Global Institute estimates the annual value of big data to US Health Care to be upwards of $300 billion (Manyika, et al., 2011). It is important that SMEs do not underestimate the potential value of their big data programs and are willing to invest capital in their implementations, as in the long term the value they can generate far outweighs the costs.

These 7 V’s should be considered in every big data venture and can be applied to all big data programs. Ensuring that each of these V’s is considered is vital in ensuring successful use of big data analytics. We now must consider where to source the huge volume of data used for big data analytics.

C. Sources of Big Data

Big data comes from a huge variety of sources, which are collated and compared in order to establish trends and correlations.

• Geolocational – Data gathered from GPS and smart devices can be used to find ‘hotspots’ where the types of customer businesses are aiming to sell to go regularly. This can aid in advertising or business location (Valentino-Devries 2011).

• Social Media – By analyzing social media feeds from sites such as Facebook and Twitter, key words may be discovered which highlight a customer’s needs for a particular product or service (McKelvey et al, 2012).

• Smart Grid – Data from various sensors as well as manufacturing data constitutes the ‘smart grid’. Data such as electricity usage can be monitored in real-time by smart meters enabling business to build an idea of when users are using the most devices, therefore are at home (Deign & Salazar, 2013).

• RFID/Barcodes – RFID (radio-frequency identification) and barcodes are present on almost every product sold today. By scanning barcodes it is very easy for a business to see which products sell particularly well during a specific period, season or promotion, enabling them to review and replicate successful sales periods (Microscan, 2011).

• Stock Feeds – By monitoring stock feed fluctuations businesses are able to predict when to buy or sell products. Stocks reflect the success or failure of products, meaning SMEs do not make unwise investments. (The Economist, 2013).

• Third Party - A number of third-party data source marketplaces have begun to appear online. These marketplaces sell datasets which can be combined with data gathered by businesses in order to identify trends in the data, making a relatively small payment for data potentially worth much more once compared to existing data and trends identified (Manyika, 2011).

While this is by no means an exhaustive list (due to the new areas of technology generating new data streams constantly), this is where the majority of big data today is sourced. Now that we know what big data is, how it works and where it comes from, we can begin to examine its benefits to SMEs.

III. APPLICATIONS OF BIG DATA ANALYTICS TO THE SME

In this section of the paper I will be discussing big data analytic strategies that can prove beneficial to SMEs. I will also be relating each case study to the appropriate “V” in order to
demonstrate the practical importance of the “V’s” model as well as identify data sources used in practice.

A. Case Study: OfficeMax

OfficeMax is a business-to-business office products retailer operating across the US, Puerto Rico, Virgin Islands and Mexico. They have taken advantage of big data analytics by tracking the company’s stock inventory with RFID and combining this knowledge with tracking of their customer service to see which products had issues arising, as well as their order system in order to see which products frequently sold together; in order to advertise combinations of products effectively. Implementing this system has allowed them to create an efficient order placement process that has reduced their operating costs considerably. Through this strategy they were able to increase the e-commerce revenue from 15%, to 25% of the total company sales. They reduced order placement costs by $400,000 annually. Finally, they gained a market share upwards of 10 percent versus their biggest competitor (IBM Corporation, 2011).

SMEs can learn from this strategy and apply it to their own big data plans. The data sources in this case study were primarily RFID tags to monitor stock levels, but also a collation of sales and order data from within the business itself. As a result the costs of data would be very low in this case study, with the value coming primarily from their interpretation and analysis of the data. The key “V” in this case study is visualization. Whilst they already had the order data and an idea of the stock levels already, the key was identifying the links between the data in the form of visualization. SME’s can emulate this cost-effectively through graph producing software (Ayasdi, 2014). Variety was also important, as with the combined data from the ordering system tracking and the stock data they were able to identify many more opportunities than they would have individually. SME’s are also able to gain several data streams for low cost, making it an effective strategy. Finally, value was important in this example; OfficeMax were aware of the potential value of using big data analytics when forming their e-commerce strategy. Although the costs of implementing a system may be daunting to SME’s it is important to be aware that the long-term gains will prove worthwhile. RFID tags are relatively cheap and an alternative which works for many businesses is barcodes, as almost every product has one (Microscan, 2011) (Smart Border Alliance, 2005). This strategy would be ideal not only for online-based SMEs but also in face-to-face retail. With barcode scanners and a big data system set up at the point of purchase both online and in-store sales can be tracked, inventory for both outlets managed and visualizations of the data produced, with meaningful insights being used in a marketing strategy to sell items frequently bought together.

B. Case Study: UPS

UPS has been using big data systems for several years, with tracking parcels and transactions starting in the 1980s. However their latest big data scheme, ORION (On-Road Integrated Optimization and Navigation) combines this tracking data with several telematics sensors. By placing these sensors in delivery vehicles they are able to see vehicle speed, direction, braking and location, enabling them construct routes based on regular delivery spots. In 2011 UPS saw savings of upwards of 8.4 million gallons of fuel through this strategy. With one mile saving per day per vehicle resulting in a saving of around $30 million the savings they are making are considerably sizable (Davenport & Dyché, 2013). UPS’s big data strategy has managed to save them 8.4 million gallons of fuel across its entire fleet of delivery vehicles, resulting in $30 million cut in in fuel costs, with potential fuel costs also slashed in air transport with more efficient flight paths and cargo management.

SMEs are able to adapt this strategy to their own needs. The primary data sources for this strategy came from the ‘smart grid’. An array of sensors produce individual data sets for the vehicle’s speed, direction and braking to judge driver’s performance in terms of fuel efficiency. A more cost-sensitive approach suited for SMEs is tracking via GPS, as many SMEs using logistics and transport already use them in their vehicles.

A key “V” involved in this strategy was volume. In order to effectively evaluate their route’s efficiency it was important for them to gather data on several weeks or even months worth of routes. For this reason this strategy will be best applied to SMEs in the logistics sector, or those who make regular journeys, as having low volumes of data may result in ineffective routes being produced. The reduced cost as a result of less fuel consumption can add up to sizable sums in the space of a year and reduce overheads drastically. This is particularly useful to SMEs, where unnecessary costs if left unchecked, can mean the difference between the success or failure of the business.

C. Case Study: Tesco

Tesco have been consistently committed to big data since 1993 with its introduction of the Tesco Clubcard. (Louis, 2002) More recently their innovation of a virtual store in Korea, where users were able to purchase items with their phones using QR (Quick Response) codes, slashed employee and premises costs. They have continued to innovate, with “scan as you shop”, a self-service scanner that enables customers to complete the entire shopping experience on their own. They have also implemented electronic shelf labels, allowing for instant and efficient price changing nationwide. These innovations have led Tesco to increase its market value from £4.7bn in 1992 up to £29.9bn in 2011 (Plant, 2013).

SMEs are able to use many of Tesco’s strategies to their advantage. The key data sources here are barcodes. They are used to not only enable purchase of products through phones, but also give customers and employees access to Tesco’s information database, in order to assist in their purchases. SME’s are able to access the barcode information for very little cost, with readers being widely available and more recently becoming available on any mobile device with a camera (CTIA, 2004). The key “V” in this situation is velocity. With the electronic pricing on shelves Tesco are able to make pricing adjustments instantly, enabling them to effectively keep prices competitive. SME’s will benefit from this by comparing their prices with other local competitors. An issue that faces many
SMEs is their inability to provide competitive pricing, but by implementing such a system this issue can be overcome. Another key “V” is volume. It is important for them to keep abreast of all their products in order to offer the entire product catalog for its services. SME’s will need to ensure that their entire catalogue is monitored and barcoded. Whilst this may come at some additional cost in the short term, long term benefits will include easier tracking of products, improved knowledge for customer service and effective pricing models. The ‘virtual store’ concept can also be used by SMEs in order to offer their product range to customers with vastly reduced running costs. Rent is significantly reduced and only minimal numbers of staff are required to run the store, which is ideal for SMEs who struggle to pay a large workforce. The result of this will be both significantly reduced overheads but also an increased novelty in the business which can lead to increased sales (Deloitte, 2011).

IV. CONCLUSION

Big data is fast becoming an indispensable business tool. However many SMEs do not currently embrace the model, primarily due to a poor understanding of big data concepts, as well as the true value that big data can bring to a business. There are also fears of extreme costs to incorporate systems and obtain and analyze multiple data sets simultaneously and efficiently. By explaining big data, clearly identifying its business potential through case studies and examining the impacts the seven “V’s” of big data can have on a big data strategy, it has been possible to demonstrate some of the key benefits of big data analytics to any SME.

Big data analytics have been proven to increase revenue in a number of businesses, reduce business overheads by ensuring efficiency of staff and services and also prove cost-effective, with some data sources already being available but not utilized and several relatively cheap or even free data stream available.

In the future big data analytic strategies may become a necessity, particularly amongst SME’s who may struggle to become established. However at this moment of time big data analytics is a way for SME’s to go above and beyond competition. Big data analytics could be the key to the success of one business against its competitors who do not take advantage of a big data scheme.

V. REFERENCES


You Should Care, To Make Them Care

Affordable Digital Marketing Solutions Using Big Data

Evrydiki Tsatsalidou
University of Derby
UOD
Derby, United Kingdom
e.tsatsalidou.1@unimail.derby.ac.uk

Abstract—the objective of this research and article is to analyze the significance of Big Data in the field of Digital Marketing nowadays. It also shows the tools of how Big Data can be used in an SME’s digital marketing strategy effectively, what big data should mean for a company, how they are actually using them and why digital marketing is a really good way to acquire data and information. The article also demonstrates how important the skilled human asset, in big data and analytics, should be to succeed in Digital Marketing and overall survive in the existing large competitiveness.

I. Introduction

In our contemporary digital life and information age, most of the people use in the biggest part of their day electronic devices to complete desired processes in easier and quicker ways. These processes have turned into online and mobile processes. Individuals can interact, communicate, share material via social networks, order food, buy products, services and book tickets online. This evolution of completing desired processes online has helped people to save a lot of time, effort and of course make value for money. (Saunders 2012)

Of course, as there is always a negative side, under these circumstances, people are getting more and more stressed as the way of living gets quicker. It is getting more and more demanding to do a number of things in a short period of time so that people can catch up on everything they want to do in a day. Tech stress is rising exponentially over the years and cannot be ignored. Tech stress can literally show and prove this unthinkable rise of data and information. (Gregoire 2013)

It is also important to mention the factors causing tech stress. To begin with people should have knowledge and familiarity with the gadgets and generally with every technological invention. In addition they should always be up to date with every new invention and be compatible with the consistent radical technological changes and improvements. This need of familiarity that every single person on earth should have is growing over the years because of the integration of technological changes in the everyday life. Every aspect of life is changing with technology like social life, shopping, jobs and their nature and so on. (Saunders 2012)

We definitely live in the information age. Information is growing with technology and there is a new term that businesses are adopting: Big Data. Big Data can be defined by the use of the three V’s volume, variety, and velocity. It is characterized by the immense amounts of data, the use of structured and unstructured data, and the speed of the data. The fourth and fifth V, veracity and value, are not generally accepted, though it is proven that Big Data indeed can provide value. (Vries 2013)

II. Big Data Can “Save” A Company

Digital marketing is the new salvation for an SME as there are many affordable online marketing solutions that can keep them active and competitive in the world of businesses, grow and expand their businesses. The benefits and advantages of these solutions can be proved and shown by a number of researches and statistics. For example, a research shows that 93% of marketers are exploiting social media to distribute their content while the impressive percentage of 71 social media users is more likely to buy a product when they are connected via a Social Network. The number of social media users is increasing exponentially over the years so the customers that are acquired by the social media are increasing, too. (Pick 2013)

Additionally, 75% of social media users “object to major companies and platforms using their personal information for commercial purposes.” And just 12% admit to having their purchases influenced by Facebook “Likes” or Google “+1s” (Hart 2013). Also, half of all social media users under age 35 follow their online friends’ product and service recommendations. (Baur 2013) Furthermore, three-fourths of marketers planned to increase strategic efforts on social media and social networking sites this year, with 68% also focusing more on SEO (Search engine optimization) and 63% on blogs. (eMarketer 2014) A whopping 95% of small businesses view blogging as an effective marketing technology tool—second only to email marketing. 15% say blogging is most effective at engaging existing customers; 11% value it more for acquiring new customers; and 69% say blogging is equally effective for both objectives. (Erickson 2013)

The fact is that businesses are not often geared up to handling the sheer volume and variety of data at their disposal, either for operational or risk management purposes. In KPMG’s recent “Going Beyond the Data” survey 96% of
respondents admitted that they did not use corporate data to its full potential, with 56% acknowledging significant benefits to doing so. Harvesting data and making it available for analytics is currently the biggest issue we face in actually putting our analytical algorithms to work. (Rossi 2014)

III. SMEs DO NOT MAKE USE OF BIG DATA FOR THEIR DIGITAL MARKETING STRATEGIES

Now, in the world of business, for those who know how to use data, things keep getting better. Companies can collect a very large number of data and information in many and various formats and with efficient ways of using them they can engage customers online using various marketing tools. (O’ Rourke 2014)

Although, there are many ways to exploit internet data aiming to integrate them efficiently into SMEs marketing, many researches and statistics show that most of the SMEs do not make use of Big Data for their digital marketing. As well, there are statistics and researches that can show and prove the huge importance and significance of using Big Data in Digital Marketing. (Pick 2013)

For Instance, statistics from the US, Canada and the UK shock showing that 85% of the consumers are searching for local businesses online, but a large number - approximately 25% - are not showing up at all in search results. Another shocking statistic from the same research states that approximately 63% of small businesses do not have a website. (Pick 2013) Without the most basic and powerful tool for online selling and marketing, small and medium sized businesses are missing out on thieving prospects from larger, web-savvy competitors. SMEs do not realize the significance and power of online marketing and they are not adopting the methods that leader companies are using to maximize their profit and generally achieve their objectives. Also a survey from Harris Interactive found that the majority of SMEs do not have a clear understanding of ‘big data’. This shows that SMEs should make more research defining what Big Data really means and figure out the significance of it. So, how big data could be used to change the promotion and the overall marketing of a business and be better? (CANDDi 2012)

IV. THE AFFORDABLE SOLUTIONS FOR SMES

While traditional marketing is still alive and effective for businesses, digital marketing is the new challenge for SMEs and it is becoming even essential especially for the start-ups who use small budget. Businesses should exploit these opportunities and use digital marketing effectively to give them the competitive advantage they want. There are several types and tools of marketing that SMEs can exploit and use to improve their overall promotion of their products and services and improve their reputation. (BusinessDictionary.com 2014)

The two main types of digital marketing are pull and push marketing. The first one refers to the marketing when the customer has an active role of seeking marketing content and in general navigating through various web pages. Usually, these actions can be taken via web searches or by opening an email, text messaging or web feed. In push digital marketing, the marketer is the one who has an active role and sends messages to the recipient so he is the one who actually “pushes” the target group to the marketing content via email, text messages and web feed. Of course both of these can be used in conjunction. (Boundless 2009)

Thereupon, there are some completely free tools online which SMEs could use. For instance, one source of broad analytics is Google analytics. This tool launched by Google is letting companies to analyze big data in the cloud. Google analytics can offer useful data for a website like how much traffic there is and how long do people stay on the website. Also, email marketing is a really effective tool if enterprises know how to use it. There is a very large number of email templates online which can be used by enterprises and as well, advice on email etiquette depending on who is the target group they are referring to. There are also big data technologies for the different data types such as email clickthroughs, conversion data from forms and browsing activity emanating from emails. (BusinessDictionary.com 2014)

Additionally, SMEs can convert customers from social media as they are proving to be essential to use for nowadays marketing. Nowadays, one wide ranged tool is of course Facebook online campaign. Facebook with ads manager offers different types of reports for the outcomes of a campaign which are showing the information businesses want. Facebook shows metrics and analytics coming from the campaign, too. (Pick 2013)

Another affordable and cost effective solution is SEO which lets a company reach the top position in the SERPS (Search Engine Results Pages) when users search for something online. Finally, there are many free demos, beta versions and affordable software online which offer numerous services. So, at least for one month SMEs can use software to try them and see if they work for their nature of business at no cost risk. (BusinessDictionary.com 2014)

V. WAYS TO ENGAGE CUSTOMERS

Interest in customer relationship management is strong. Across a wide spectrum of industries, companies have come to realize that their customers are central to their business and that customer information is one of their key assets. (Berry et al 2004) Two main things that the business should make to succeed this are provide satisfaction in the side of customer and customer rewarding. This will result in for the product and service to have value for the customer and increase the loyalty of the customer. ‘Increased customer loyalty is the single most important driver of long-term financial performance’ Dave Illingworth, Lexus US (Doyle 2007)

There are many ways to increase the market share and attract new customers by using big data in SMEs. SMEs should be aware of this and use the online networking opportunities and try to exploit them to have an increasing number of more satisfied customers and a better fame and resonance as a business. In summary, the ways to succeed this is corresponding to the success of the transformation of data collected and transform them into useful information for the enterprise. The affordable tactics for an SME and a startup
business would be to carry out Online Surveys to understand customer needs (Time, money saving), have Online Customer Service, for instance, real-time chat with an assistant to solve the customer’s problem, Direct Phone Service to solve the customer’s problem, Collecting Online data from mobile users (smartphone applications, tablets), Home Users (Desktop PCs), Web 2.0, Mobile optimized websites, using Social Media to Administer Online Surveys, SEO, Personalization, automation, CRO (Conversion rates optimization web design), Redirection and On click paid Ads. (BusinessDictionary.com 2014)

Better digital marketing means general improvement of a business performance, overall better marketing, bigger customer satisfaction, retention and engagement and a more competitive company. (Berry et al 2004)

VI. ADVERTISING MEDIA

A. The Changing Nature Of Advertising

Each day consumers are bombarded with advertising messages for all kinds of products through many different advertising media, ranging from TV to direct mail. Traditional mass marketing used a small group of advertisements on a limited number of media to reach as many people as possible. Developments in alternative media channels and the use of customer databases have enabled marketers to create individualised advertisements based on the customer’s behaviour and preferences. While traditional media such as TV are still important, a whole range of other options such as Internet, DVDs and interactive television are becoming available. This has caused marketing departments to explore ways of developing customised statements and small press runs, highly tailored to the customer. The variety of different options open to marketers will increase dramatically and marketing departments will have to develop new capabilities in these areas. One of the biggest trends in advertising is the fragmentation of media. Marketers for FMCG (Fast-Moving Consumer Goods) goods can no longer rely on the effectiveness of large advertising budgets which promote their products on TV and the job of reaching customers has become much more complex. (Bickerton et al 1996)

B. Developing An Advertising Campaign

An advertising campaign consists of designing a series of advertisements and placing them in various advertising media to reach a particular target market. In general, the following guidelines can be used to develop an advertising campaign for most types of organization: identify and analyse advertising target market, define advertising objectives, determine advertising budget, develop media plan, create advertising messages, execute campaign and evaluate advertising effectiveness. (Bickerton et al 1996)

VII. BIG DATA IN IMPLEMENTATION

A. The People Can Manage Big Data

It’s the people in a business who will make the real difference: training one person to use web analytics tools, understand and interpret collected data can allow a small business to begin allocating resources more efficiently very quickly. Data from the firm’s website can pinpoint problems such as uninspiring marketing messages, ineffective product campaigns, and help the company target an audience most likely to buy. Well worth the investment. “Small businesses need all the help they can get to attract inbound enquiries and make the most effective business decisions today. The web can no longer be ignored as a source of information and marketing to engage potential prospects, especially as more and more of the big players embrace sophisticated web analytics with big budgets.” (CANDDi 2012)

The companies should hire the right stuff to manage Big Data. Of course, each company has its own individual strategy for hiring the staff to manage Big Data for the digital marketing but there are main skills that they should have. The employers should have a high level of creativity, have an understanding of workforce analytics, data science skills, analytical skills and technical skills in some kind of technical background. The process of hiring should not take a lot of time, money and effort as now many organizations are restructuring and thinking through hybrids of current positions or creating entirely new positions within their structure to effectively manage, analyze, and put their Big Data assets to good use. Where is an organization in terms of personnel? Does it have anyone on staff with analytics knowledge or expertise? (Barker 2014)

Someone within the organization needs to have working knowledge of analytics, its advantages, benefits, and best use cases. If the CIO (Chief information officer) or CMO (Chief marketing officer) doesn’t have expertise in analytics, it is indeed needed to consider hiring and investing in someone who does. MGI (Mouse Genome Informatics) calls analytics “the world’s hottest market for advanced skills” and the leading organizations are snapping up personnel with these skills sets, making them expensive to hire and in high demand. Another viable option to consider is training the key C-suite level staff in analytics beyond the basics. (Waddell 2009)

VIII. CONCLUSION

Overall, there are many digital affordable solutions for SMEs so that they can use Big Data. This research shows and proves that SMEs in general do not make use of Big Data in their overall digital marketing strategy. Needless to say, they also do not make right use of digital marketing tools, which are requisite nowadays. (Pick 2013) The outputs, the data and the information, from the use of digital marketing should be used by businesses. Firstly, they should research and realize the significance of using Big Data nowadays and find out what leader companies do to succeed in their sector. Also, SMEs should constantly exploit opportunities that are presented to them and make the most out of them to achieve their desired results. What is generally hidden behind the success of every business is that they are effectively using large amount of data and information. With the amazing range of tools available online the outputs of digital marketing can turn into valuable information. It is a very important part for each company to figure out what kind of data are relevant, could be helpful and play a major role in the business’s aims. (Vries 2013) Thereupon, this article shows the changing nature of
advertising and the steps that should be made to develop an advertising campaign. (Bickerton et al 1996)

Lastly, it can be concluded that the process of analyzing data is not just a software’s work. The human resources of the company should be trained to use analytics tools and depending on the nature and sector of the business each company should hire the right staff to manage Big Data. (Barker 2014) There are some main skills that every Big Data manager should have, like creativity and technical skills, to manage data. (Waddell 2009)

IX. REFERENCES


Doyle, P. (2008) Customer Loyalty - Value-Based Marketing. 2nd Ed. West Sussex: John Wiley and Sons Ltd. p.73


Do SMEs Benefit More From Big Data Or Small Data?

Comparing The Training, Technologies And Cost Considerations Of Each

Yasmin Walji
School of Computing and Mathematics
University of Derby
Derby, Derbyshire
y.walji1@unimail.derby.ac.uk

Abstract—With less than 0.2% of SMEs engaging with Big Data, questions are raised on how to either enhance the knowledge of SMEs in Big Data or find a different solution to improve aspects of their marketability. Small Data could be more appropriate for SMEs as it deals with specific problems and provides actionable data rather than collecting a massive amount of data, especially where that data may not exist at their business scale. However, with each type of data analysis come many different considerations that need to be made on training, technologies and cost, all of which could influence which type of data analysis is most beneficial to individual SMEs.

Index Terms—Big Data, Small Data, SMEs, training, technologies, cost

I. BACKGROUND

In 2010, the Big Data industry was worth more than 100 billion dollars and has grown by greater than 10% each year. (The Economist, 2010) As of 2012, more than 2.5 Exabytes of data is being created each day (IBM.com, 2014) and it is not surprising that large corporations, such as Apple and Google, collect the majority of that data. (Davidson, 2014) A lot of resources need to go in to collecting and analysing the data which is why in 2012 less than 0.2% of small to medium enterprises (SMEs) invested in Big Data. (SAS UK & Ireland and e-skills UK, 2012) Instead, it can be argued that they benefit more from targeting certain types of data which provide answers for more specific needs than the broad trends which usually come from Big Data. (Grabova and Darmont et al., 2010) The term “Small Data” has been given to data which come from focused areas such as web traffic and surveys.

All businesses can benefit from collecting data to enhance their services and reach, but SMEs are overwhelmed by Big Data. In 2012, 22% of SMEs had “poor” or “very poor” knowledge of the related concepts and technologies. (SAS UK & Ireland and e-skills UK, 2012) They also may not require the sheer quantity gained from Big Data, especially if their services can only cater to their local area. For example, a small business is not going to be concerned with the interests and desired services of international consumers if they are only able to cater to their local city, county or country.

This paper will compare the resources required for both Big and Small Data, and how they can have an impact on SMEs. Firstly, the concepts need comparing and given context.

II. BIG DATA

Datasets that are highly detailed and vast in quantity are known as Big Data. They come from many sources, often in real time, such as public clouds, social media content (blog posts, status updates, check-ins, etc.), transactions, web activity, and so on, and require networks of machines to store and process all of the information. (Russom, 2011) The purpose of collecting these huge volumes of data is to establish trends by making connections between different sets of data.

Businesses that are national or international can benefit from establishing these trends to distinguish differences and similarities in sales, satisfaction, interaction, etc. so that they can enhance their services to local needs. They can also gain greater reach to potential consumers, especially through social media where most of their opinion-based data is likely to come from.

III. SMALL DATA

Unlike Big Data, Small Data is not recognised as widely as a concept, but rather it refers to a collection of methods which acquire more organised and actionable data. (Bonde, 2013) A simplified definition of Small Data is a focused dataset that can be processed on a single machine. (Pollock, 2014) Things like e-mails, surveys, social media data mining, etc. which are aimed at specific groups or keywords rather than collecting any and all related data.

Businesses that are local or national with only a few branches may have smaller current and potential customer bases which they can survey to establish their interest and needs in their service. They can also collect data on their web traffic to understand where the customers have come from, to aid in targeting other areas to reach more people. The data that is collected is more focused on answering specific questions than establishing broad trends, although sometimes trends are desirable, such as seasonal or local differences.
IV. INVESTIGATION

Aside from what a business needs in terms of amount and type of data, they must also consider the resources they are able and willing to invest in. This section will discuss the training and technologies for each type of data analysis and their potential costs, followed by specific advantages of different types of Small Data to highlight the kinds of areas businesses may want to improve (such as profitability, reach, productivity, etc.) and which type of data analysis will be most beneficial to them, outside of the training, technology and costs.

A. Training

For Big Data, teams which are capable of dealing with high volumes of complicated data are necessary, regardless of whether the processing is done in-house or externally. People who are already trained can be hired, but if the business already has an IT team, they may deem it more suitable to invest in their training. The employees that are already skilled at developing enterprise data driven applications with languages such as Java, Ruby and Python will find it easier to train in Big Data. (Badcock, 2013) Therefore, if the company already has a developing team, hiring an expert in Big Data to guide and train within the team will be more beneficial than getting each employee to take a course or individually learn. Additionally, for in-house networks, the IT team need to be able to provide continual support for the hardware and networking of the machines.

If the business does decide to invest in courses to train their employees, these range from broad courses that cover Big Data as a whole, (Executive.mit.edu, 2014) to more focused courses which deal with certain types of data collection and analysis. (Econsultancy, 2014) A lot of these can be done online which means employers can decide if they would impact on the current flow of the business. Instead, they may expect employees to dedicate their own time outside of working hours to complete these courses.

These focused courses also apply to Small Data as employees can train in areas which are most relevant to the collection methods and how the data are being analysed.

The downside to all of this training is that it can be seen as never ending as technology is ever changing and expanding. (Bughin and Chui et al., 2010) To truly stay at the top of a business’s respective market, they must have a grasp on the latest technologies and techniques, which means many staff changes or continuous training. However, hiring someone who is skilled in many developing areas will mean they are better equipped to learning and adapting their knowledge for data analysis. (Badcock, 2013) and similarly for those who are skilled with IT support will be able to keep up to date with the latest hardware and networking advancements.

B. Technologies

In terms of hardware, as mentioned in the Background section, Big Data requires a network of machines. This is mainly for the processing and storing as it takes a lot of processing power to deal with such high volumes of data, and of course the high volumes of data need to be stored somewhere. In contrast, Small Data can be processed on a single machine, which does not have to be top of the line, but still have ample storage. Of course, a few machines in a small network rather than something like a datacentre may also be appropriate if a lot of small data is being dealt with rather than just one or two types (i.e. if a business is transitioning from small to big).

Commonly used software for Big Data analysis is some form of MapReduce, such as Apache Hadoop. This software splits the data equally among each node in the network assuring that each part gets processed as efficiently as possible. (Dean and Ghemawat, 2008)

The Big Data industry is also at the forefront of the move towards NoSQL databases, such as Apache Cassandra and MongoDB, as they are better suited for clusters of different types of data, rather than being forced into a data model. (Datastax, 2013)

A lot of tools for collecting and analysing Small Data can now be used through websites. For surveying, a business can launch the survey from their own website or e-mail hosts, and either use a web service like MarketSight to analyse the results, (Marketsight.com, 2013) or do it themselves in a spreadsheet tool like Microsoft Excel. Similarly, for web traffic and transaction information, using some JavaScript code on their website they can capture visitor information themselves, or instead have it sent on to dedicated services like Google or Yahoo Analytics. (Dubois, 2010)

For analysing a business’s social media reach, web services like Twitalyzer can provide information on interactions and impact, based on sharing and replies.

C. External services

If businesses do not have the resources to manage the networks or analysis themselves, there are many services to aid them. They can provide cloud storage so businesses don’t have to have their own network of machines and staff to manage them. They can also do everything from collection to analysis. It’s just then up to the business’s staff to deal with the information it provides.

Because of the costs of having large networks of machines or storage (which will be covered in the next section) as well as their maintenance, a lot of businesses that invest in Big Data outsource the processing and storage. Large corporations like IBM and Microsoft which have vast datacentres all over the world offer services to store, process and analyse data, meaning businesses which cannot invest the resources to hiring/training staff and purchasing the hardware and software are still able to benefit from data analysis without getting too far into the technical side. (Sallam and Richardson et al., 2011)

Some of the services mentioned in the previous section for Small Data were external services, such as MarketSight, Google Analytics and Twitalyzer. Usually these are more for aiding businesses where they may not be able to have fully trained employees, but rather people who are competent enough with a computer and can learn to use these services instead. (Dubois, 2010)
D. Cost

Prices vary between providers and countries, so this section is not going to go into specifics for each service or employee, but instead act as a guide as to whether a business would be looking at spending hundreds or thousands of pounds for certain services.

To hire someone already qualified in Data Analysis, whether they have any other IT skills or not, their annual salary could be anything from £40k-£100k+. Those that are skilled in Hadoop development could be looking for anything above £90k, while general “Big Data”/“Cloud” Engineers will look for £50k - £70k per annum. (Badcock, 2013)

To train existing employees instead, courses that cover Big Data as a whole can start anywhere from £1,500. (Executive.mit.edu, 2014) For more focused courses, such as in Web Analytics, prices range from £500-700. (Econsultancy, 2014)

Services like Google Analytics are free, unless a business wishes to have more extensive guarantees and support at a fixed annual fee of $150k (approximately £90k). (Google.com, 2014)

To outsource everything to a company like IBM, businesses can get a bespoke quote, meaning they will be paying for what they need rather than set packages. (IBM.com, 2014)

Hardware costs vary between brands and specs, but where storage is concerned, hard drives are increasing in size and becoming more affordable. External hard drives are also a lot more affordable and make it easier to add more space or create backups. A typical 1 Terabyte external hard drive on average starts at £55. (Idealo.co.uk, 2014)

V. ADVANTAGES OF SMALL DATA

Before fully considering any of the above, a business needs to recognise their desired outcome. Each type of data analysis has its own advantages to a business. Some are more suited to increasing their customer base or reaching to other types of customers, while others improve productivity or efficiency of the business itself.

Big data aims to improve most or all areas of the business by establishing trends which influence where changes need to be made, whereas Small Data targets certain problem areas or areas where there is room/desire for improvement. Of course, within Big Data are Small Data methods which help tackle specific areas/questions.

Web Analytics aims to increase profitability by enhancing the customer’s experience on the business’s website. (Waisberg and Kaushik, 2009) The goal is to attract the right people and persuade them to purchase the business’s services. Web analytics can tell them where their visitors are coming from and highlight how popular certain aspects of their website are. This helps them to aim the most important or appealing aspects of their website to the areas where most visitors pay attention.

Surveys which ask the right questions can highlight many different areas for improvement, particularly to improve customer satisfaction, interaction, profitability, etc. (Penn State University, 2006). Similarly, by using social media data mining, businesses can see what is being shared about them to gauge current satisfaction levels or needs for improvement, especially in how to extend their reach.

As mentioned, even businesses that have invested in Big Data need to consider Small Data solutions for analysing particular data. The data and trends are usually not enough to answer specific questions, like what new areas a business can explore, what customers really think of them, how to promote themselves better to reach more people, etc. (Proffitt, 2012) They also will not have any answers for where their data has not breached, such as how to expand their business nationally or internationally if their data is currently only concerned locally. Initial surveys and data mining may need to be carried out to gauge potential areas for data collection and answering questions they may not have even thought of yet.

VI. DISADVANTAGES OF SMALL DATA

The main problems with Small Data are dealing with the veracity and validity of the data. (Normandeau, 2013) When you have larger datasets, it is easier to make a generalisation as the more data you have, the more truth you can gauge from it. In smaller quantities, there is less representation and more room for erroneous/false data.

A simple analogy is flipping a coin. It is clear there is a 50% chance the coin will land as either heads or tails, however when flipping the coin only 10 times, it may land on heads 7 times. If you increase the amount of times to 100, it may land on heads 60 times. If it is increased to 1000, it may land on heads 525 times. The larger the sample size, the closer to a truer representation the data becomes. (Walker, 2013)

Additionally, if the coin is biased, the larger the sample size, the more evident it will be there is something wrong.

There’s also the variety of data, especially when obtained from social media. SMEs need to invest further in working out how to deal with the different types of data, such as general text, check-ins, statuses, comments, tags, pictures, videos, etc. and any combinations of them. Big Data is all about variety in large volumes, so considerations are already made with the processing power, software, and training. (Joseph, 2012) When dealing with just social media in Small Data, although there may not necessarily be the volume, dependant on their business scale, there will still be the variety which some SMEs may be overwhelmed with.

There is also the problem of reach in social media, particularly on sites such as Facebook where they have implemented paid promotions to boost reach for statuses and ‘Like’ pages. (Naidu, 2014) SMEs may be overshadowed and have a skewed sense of their actual/potential reach because larger brands and companies can afford to promote themselves. This can also affect the potential data in general regarding the opinions on those they have come into contact with who may not be aware of their social media presence because it is not spread widely enough.

VII. CONCLUSION

There are a lot of different considerations for businesses wishing to take advantage of data analytics, especially those with limited resources. A majority of SMEs have not yet
immersed themselves into the world of Big Data for various reasons, most of which stem back to not having the required knowledge or even the need for vast data collection.

For SMEs which are on the smallest end of the scale, using free web tools which are catered for less technically skilled people are more appropriate than hiring a specialised team to set up their own collection and analysis of the data. For SMEs on the larger end of the scale, a dedicated team can be hired, although it still is not necessary for them to be specialised to begin with as training them in relevant courses should be more affordable. They also do not necessarily need to be proficient in IT and networking support as, depending on the amount of Small Data being processed and stored, a single machine should be sufficient.

As there are many different types of data analysis to choose from, SMEs benefit much more from Small Data as they can pick and choose which specific areas they are going to analyse, as opposed to collecting a lot of information where not all of it is actionable or relevant to their current needs/access to resources. However, if SMEs are trying to get the most reach online, particularly in social media, they need to be prepared for the vast variety of data they will be collecting, and be sufficiently trained or have the right software to analyse it. This in itself can be a large restriction, but there are short and relatively cheap online courses that are focused on specific areas which an SME may wish to invest in if they feel a social media presence will have a great impact on their business.

VIII. REFERENCES


REFERENCE LIST


223


Gov.uk. (2014) £73 Million To Improve Access To Data And Drive Innovation. [Online] Available from:


Michael, J., Berry and Gordon, S. (2004) Interest in customer relationship management is strong, Data Mining techniques for...
marketing sales and customer relationship management. 2nd Ed. US: Wiley publishing Inc. p.383.


INDEX OF AUTHORS

Mohammed Saeed S. H. Abdulla.................................................................7
Farah Ahmed .............................................................................................12
Mohammed Al Yousif ................................................................................16
Awfa G S Al-Adawi ..................................................................................20
Hassan Al-Emadi ......................................................................................24
Gareth James Allen ..................................................................................28
Maha Alshahri ............................................................................................33
Adel Alsooj ................................................................................................38
Ali Al-suwaidi ............................................................................................42
Mathew Bateman ......................................................................................47
William Briggs .........................................................................................52
Adam charlton ...........................................................................................57
Alex Clark .................................................................................................61
Marcin Tomasz Drozdz .............................................................................66
Oliver Fox ..................................................................................................71
Mohammed Haroon ..................................................................................75
Robert Heeley ...........................................................................................80
J. Hill ...........................................................................................................85
James Adam Hunt .....................................................................................90
Edidiong Emmanuel Inokotong (Student no: 100323055) .........................94
Samuel James ...........................................................................................99
Michael King ...........................................................................................103
Vasiliki Kozi .............................................................................................110
Daniella Kypri ..........................................................................................114
Evaldas Luksys .........................................................................................119
Mawhinney, Liam ....................................................................................125
Amrith Nagra ...........................................................................................130
Umair Farooq Naru ..................................................................................135
Feyisayo Obisesan ...................................................................................140
Marialena Panagiotidou ..........................................................................145
Gwyn Dafydd Owen Perkins ....................................................................150
Simon Ranson ..........................................................................................156
Joe Rawlings ............................................................................................160
Joao Vasco Zurzica Reis ........................................................................164
Tim Ride ....................................................................................................170
Stephen Ridgway .....................................................................................174
Kareem Samarah .....................................................................................178
Tomas Satala ............................................................................................182
Umar Shakil .............................................................................................187
Diana Silva Caires .................................................................................................................. 192
Wesley Simms .......................................................................................................................... 198
Matthew Smith .......................................................................................................................... 203
Josh Trow ................................................................................................................................... 208
Evrydiki Tsatsalidou ............................................................................................................... 213
Yasmin Walji .................................................................................................................................. 218
END OF BOOK