sine
CONSULTING

Data Analytics Report

Siddharth Shrivastav

Under the guidance of Prof. Gerry Conyngham

SECTION A

What is Data Analytics?

Data is everywhere in the world and is expanding at an alarming rate in the world, and it is said that 90% of the data is created in the past few years (Wall, 2014). According to (IBM, 2012), 2.5 quintillion bytes of data or to be specific 2.5 billion gigabytes was produced every day in 2012, wherein the data comes from a lot of different sources. The sources are like weather climate information, posts on the social media websites, digital pictures, online videos, transaction records in the banks, mobile phone records (IBM, 2012) and amongst them it is interesting to know that 75% of the data are unstructured (Wall, 2014). As the mobile phone penetration is about to increase to 70% in 2017 from 61% in 2013 (Wall, 2014), the data is about to about to grow drastically which is simplified as 'Big Data'. Big data can be referred as the datasets which are beyond the control of the standard software management tools to capture, store, manage and analyse (Manyika, et al., 2011). The are some provisions being made to store this data (Wall, 2014), but the real problem comes to manage these data and there comes the role of Big Data Analytics. According to the computer giant IBM, Big data analytics is the usage of advanced analytic techniques against the vast and diverse dataset (IBM, 2017). Big data and data analytics is used to define the data sets and analytical methods respectively that are broad and complex which requires different and advanced data storage, analysis, management and visualisation technologies (Chen, et al., 2012). The process of collecting, organising and analysing large data sets (called big data) to discover patterns and other useful information is called Big data analytics (Beal, 2016). Even though the framing of words may be different within all these definitions given by various people but the idea behind all is the same which clearly denotes that the Big Data Analytics is about managing the data efficiently by the used of advanced analytical techniques (Russom, 2011). The big data analytics include managing datasets which are structured, unstructured, streaming data and batch data and also data of different sizes from terabytes to zettabytes (IBM, 2017). Big data analytics is vital in today's world as there is loads of data, but it is also important from the organisational perspective which is as follows,

- Big data analytics helps the organisation to harness their data, and it is also used to identify new opportunities which in turn leads to smarter business decisions, higher profits, efficient operations and more happier customer (Wagner, 2014).
- Focussed on the data which is extremely vital for the firm by obtaining better data context, For example, Data which is available on the social media like Facebook, Instagram, Snap Chat, Pinterest, etc. are an ideal way to understand today's customer. So, these raw data from social media will be aligned in a way to make business sense which helps managers in the decision-making process (Wagner, 2014). A sound and thoughtful approach big data analytics strategy ultimately make organisations smarter and more efficient.

- Business Intelligence can be benefitted from big data analytics as it helps in more accurate business insights, an understanding of the business change, planning & forecasting and the identification of the root cause of the expenses (Russom, 2011).
- Analytical applications can also be beneficiaries of the big data analytics as it helps applications to detect frauds, quantification of risks, market sentiment trending and also helps in automating decision-making process such as approval of loans (Russom, 2011).

By now we have understood that what is big data, big data analytics and the need for data analytics but now we will acquire insights into the tools and techniques for big data analytics.

Tools and Techniques in Data Analytics

Let us first understand the classification of data before moving towards tools and techniques. Data is broadly classified into Structured data & Unstructured data (Conyngham, 2017) and both are used extensively in big data analysis (Schaefer, 2016). Data which is highly organised as information and systematically drawn in tables, in fixed fields and easy to detect with the help of search operations and algorithms are Structured data (Schaefer, 2016). On the contrast, Data which is derived from multiple and complex sources like website blogs, social media content, multimedia content, email content, customer service interactions and sales automation are called Unstructured data (Schaefer, 2016). Tools for managing and analysing structured data are,

- Spreadsheet Packages – Spreadsheets is an intuitive user interface which helps users to compare, contrast, interpret, manipulate the data and a flexible data model with the ability to add rows, columns, or tuples seamlessly (Bendre, et al., 2015). The best part of these spreadsheets is that it is used by non-programmers including consultants, finance professional, statisticians and physical scientists (Bendre, et al., 2015). Some spreadsheets which have found pervasive utilisation in the ad-hoc tabular data analysis are VisiCalc, Lotus 1, Lotus 2, Lotus 3, Microsoft Excel and Google Sheets.
- Relational Database Management System (RDBMS) – A relational database management system is based on the model invented by Edgar F Codd, the IBM's British Scientist (Management Study Guide, 2017). RDBMS is a structured query language used to create, update, and administer a relational database (structured data storage and efficient retrieval of information) (Chen, et al., 2012). Some examples of RDBMS are Oracle, MS SQL Server, IBM's DB2, My SQL and Microsoft Access (Conyngham, 2017).
- Statistical Software – Statistical software such as SPSS, SAS, MINITAB and STATA are the specialised computer program designed to analyse structured (predominantly quantitative data or categorical) dataset efficiently (Conyngham, 2017).
- Online Analytical Platforms (OLAP) – The technology behind the business intelligence is Online Analytical Platforms (OLAP), which is used for data discovery, trend analysis,

capabilities for unlimited report viewing, business performance management, financial reporting, simulation models, complex analytical calculations and predictive planning (OLAP.com, 2017).

Due to nature, volume and source of unstructured data, it requires another tools and techniques which are efficient to manage them. Even though there are many options for the big data analytics which is worthy, but the problem is, it is hard to know all and select the best one as every tools and techniques have its own pros and con. The most important part of big data analytics is to advance the analytical tools such as predictive analytics, statistical analysis, complex SQL, data mining, data visualisation, natural language processing, artificial intelligence and database methods that support analytics (Russom, 2011). Some of the big data analytics which is trending in the market are,

- Big Database tools
  - Hadoop Distributed File System (HDFS) – Hadoop is an open source framework for software for storing the data and running applications on the clusters of commodity hardware (SAS, 2017). Although the interest of the users in Hadoop is very high due to sceptical data and file system could be a good fit for the diversity but, they are rarely adopted as there are complex data types associated with it like weblogs and XML documents (Russom, 2011).
  - MapReduce – MapReduce is the distributed file system where a user defines a data operation such as query or analysis and uses the platform 'Maps' across all connected nodes for distributed parallel processing and data collection (Russom, 2011). The mapping and analytic processing work in spite of disparate data types scattered across many distributed files and then the platform consolidates and reduces the responses that come back (Russom, 2011).
  - Clouds Based Solutions – According to IBM, Cloud based solutions, are the software as a service that is owned & operated by others that connect data of computers users with the Internet which allows the organisation to combine data from various sources and different communication channels (Khanna, 2016). The preference of people to the cloud is mainly due to the fear over the data security and governance (Russom, 2011).
- Data Visualisation Tools - The best options for big data analytics and a natural fit is advanced data visualisation (ADV) which are used to represent millions of several data point and have the ability to present that data structure onto the computer screens (Russom, 2011). Visualisation approaches are the techniques used to create tables, images, diagrams and other innate display in ways to understand complex big data (Chen & Zhang, 2014). For instance, Tableau and Oracle DV (Conyngham, 2017).
- Analytical Tools

- Web Analytics – Web Analytics is an approach where web metrics are extensively collected and analysed that may meet the organisational demand for an efficient evaluation of online strategies by understanding the relationship between the customer and the website (Phippen, et al., 2004).
- Text Analytics – Text Analytics is mainly pooling of information from text sources which can be used for summarization, sentiment analysis, explication, investigation and classification of the data (Gartner., 2017).
- Mobile Analytics – Data which is gathered, tracked and reported from internet sites which are accessed from an unprecedented number of mobile devices users to understand the behaviour of the consumer (Darden, 2016).
- Social Network Analytics – Social Network Analysis has emerged as an essential technique in today's world which views social relationships in term of network theory which consist of nodes and ties (Chen & Zhang, 2014).

- Optimisation Methods – Optimisation Methods have been applied to solve quantitative problems in the field of physics, biology, engineering, and economics wherein several computational strategies for addressing global optimisation problems are discussed and therefore can be highly efficient (Chen & Zhang, 2014).
- Data Mining – Data mining is a set of techniques which includes clustering analysis, regression, classification and association of data to extract valuable information from it, which involves the methods like machine learning and statistical programming language (Chen & Zhang, 2014).
  - Machine Learning – Machine Learning is an essential subjection of artificial intelligence which is aimed at designing of algorithms that allow the computers to evolve behaviours based on empirical data (Chen & Zhang, 2014).
  - Statistics – Statistics is the science of collecting, organising and interpreting the data which is used to exploit correlations and relationships between different objectives (Chen & Zhang, 2014). One of the best statistical programming language used is R Language (Conyngham, 2017).

Opportunities and Growth provided by Big Data & Data Analytics to Organisation

Data has always been a source of competitive advantage and business value, but now the efficient use of data has added a new dimension, as 81% of the respondents in the report of EY agree that data is the heart of decision-making process (EY & Nimbus Ninety, 2015). In the current scenario, Big data is completely changing the way the businesses take a decision, compete and operate in the market wherein companies that invest in data and derive value from it will have a discrete lead over its competitors (EY, 2014). But, some organisation have not understood the importance of big data and have ignored the real potentials of it. The reports of (EY & Nimbus Ninety, 2015),

shows that 81% of the businesses understand the importance of big data in improving efficiency and business performance. It also helps in understanding the preference of the customer in a better way and controls needed to support value-driven decision making (EY, 2014). On the same note, reports by (Deloitte., 2013), says one of the reasons for not using analytics strategy for organisation's strategy is the lack of centralised approach to capture and analyse data for company's use. Further, in this section, we will have a thoughtful approach to the scenarios of growth and opportunities provided by the big data analytics to the organisations which are as follows,

- Due to the growth of social media, widespread use of sensor technology, website blogs and other factors there is the major increase in the creation and availability of data which can provide organisations with valuable insights regarding customers, resource allocation, customer preferences, behavioural trends, production processes and supply chains (Forfas, 2014).
- As the economic returns from the use of big data are extremely high wherein a research of 2012 shows that restitution on investment in big data of more than 200%, the demand for big data analytics is increasing day by day and also the demand for individuals who facilitate data exploitation will increase (Forfas, 2014).
- Big data analytics can also drive business performance by enabling agile planning, more accurate forecasting and better budgeting (Forfas, 2014).
- Regarding competitiveness and capturing potential value in the market, companies should take big data seriously as the usage of big data and data analytics will be a key basis of competition and also help in the growth of individual firms (Manyika, et al., 2011).
- Big data and data analytics will create opportunities for the companies offering consulting services by influencing the market in the most efficient manner, and also large organisations are willing to deploy in-house useful analytical strategy and creating opportunities (EY, 2013).
- Big data and data analytics can also be very helpful in monitoring remote services within specific industries (EY, 2013).
- Talking from the perspective of the government and public sector organisation, big data and data analytics can do wonders by securing the citizen information, even patients based insights and confidential data which can be used negatively against them (EY, 2013).
- Data Analytics gives policymakers the ability to test potential solutions in advance which may not be perfect, but they signify a more transparent approach to predict whether a policy that worked in one country will be valid in another which will improve their performance in responding to needs of the society (Barbero, et al., 2016). United States has taken the initiative to take the support of big data and data from other federal departments and agencies to address the legislative issue for decision making and for media announcements (Batarseh, et al., 2017).

Role of Data Analytics in Large Organisation and SME's

It is interesting to know how big businesses are using data and analytics to take a strategic and operational decision by deriving meaningful insights from the data and converting the knowledge into action. But it is hard to digest the fact that only 1100 firms or 0.1% of the total 2.1 million SME's in the United Kingdom have employed data analytics whereas, in the large organisation 14% have used data analytics (SAS & e-skills UK, 2013). In both the cases, if SME's are compared to large-scale firms, even though the application in a large organisation is high but it is not satisfactory. SMEs are significant players in the analytics market and nurturing their enhanced use of these techniques is essential for future investment and innovation (Forfas, 2014). Whether it be large scale organisation or SME's, Data analytics has helped various sectors like healthcare, travel & hospitality, retail, governmental body and public institution in some or the other way (SAS, 2017). Let us first analyse the role of data analytics in small and medium scale enterprise and then move on to large organisation.

- Role of data analytic in SME's
  - Data analytics helps SME's in analysing super brands that are attempting to work out on what people are saying in regards to them over an assortment of unstructured information areas, for example, email and social networking platforms. Most small and medium organisation will discover what they require in existing data mining instruments (Sena, et al., 2016).
  - Data analytics helps SME's to keep an eye on smaller data sets from CRM platforms, social media or email marketing programmes that can provide much-needed insights to help them in establishing the strategy for their businesses by understanding customer behaviours and showcase patterns (MacInnes, 2016).
  - As SMEs make and store more value-based information in digital form, they can gather more precise and point by point performance information on everything from item inventories to sick days, and in this way uncover variability and boost performance. Big Data can also support alliance in SMEs by making ongoing answers for difficulties in each industry which can be accomplished by using the openness for primary leadership and decision making (Sena, et al., 2016).
- Role of analytics in large organisation
  - Big data analytics ability enables firms to coordinate different information sources with moderately little exertion in a short frame of time combined with low cost of storage which empowers associations to built consolidated analytics for decision-making process (EY, 2014).

- Big data innovations help enterprises from the general accuracy and cost challenge by allowing them to store information at the most reduced level of detail, holding all information history under reasonable expenses and with less effort (EY, 2014).
- Big data analytics helps the organisation in deploying data and information, storing, processing which gives control over a grid of commodity hardware, with unconstrained adaptability and flexibility to adjust to changing information landscape (EY, 2014).

Recommendations

It is vital for the organisations to explore big data so that they can know and discover business facts that companies never knew, as well as the opportunities for the new customer segments and cost reductions. It is recommended that organisations think big data as a chance to grow and not a problem, as even though big data management presents technical challenges but it can lead to cost reductions and uplift the revenue (Russom, 2011). It is important for the organisation to be updated on the theme of advanced data analytics as it is a collection of tools and techniques that help organisations in the long run. Some of the advanced data analytics methods are predictive analytics, data mining, statistical analysis, complex SQL, data visualisation, artificial intelligence, natural language processing, and database methods that support analytics (Russom, 2011). In the end, I want to conclude by saying that, companies have to be aware of the barriers to data analytics and ways to overcome that as it can be a hurdle for the firm in the long run.

SECTION B

Recommendations of Tools and techniques by SINE Consulting

As we have seen various tools and techniques in section A, we as a consulting firm would like to recommend Tableau Software and Python Language as a new Data Analytics and Visualisation techniques. Below are the details of these techniques as follows,

Tableau

Tableau is a visual analytics platform, it is extraordinarily centred around making this solely visual interface (Butler, 2016) which is based on the data visualisation science and it provides unique analytical experience for the end user. Being a Business Intelligence (BI) tool, it helps in business analyzation of the big data by solving complex business related problem. Any kind of data source can be associated with Tableau virtually, be it corporate information distribution centre, Microsoft Excel or online information. It gives clients quick bits of knowledge by changing their information into delightful, spontaneous representations within a couple of seconds (Interworks, 2017). With Tableau, it becomes easier to use the strength of database and associated database efficiently which helps you upgrade the inquiry execution by transforming and loading highlights. Changes can be made in information sorts, connect, split, join, mix

information, make sets, containers, bunches, etc. which implies you don't really need to contribute on a different arrangement (Adhikari, 2017). Tableau has proved to be more famous among the business analytical tools till now because of its unmatched nature of making things simple and easier. Information in Tableau is simple to share and utilise regardless of what the requisites are (Adhikari, 2017). Tableau online is cloud-based business intelligence gadget that can empower any association to impart perceptions made in Tableau Desktop to other individuals (Butler, 2016).

Python

Python is a highly translated and arranged, a programming language with dynamic semantics, open sources, flexible, powerful and easy to use (Shaikh, 2016). Python's high level built in information structures joined with compelling writing and vibrant commands, make it incredibly appealing for Rapid Application Advancement and Development, and for use as a scripting language to associate existing parts together (Pointer, 2016). Python is straightforward, simple to learn punctuation underlines syntax and in this way decreases the cost of program maintenance. Python underpins modules and bundles, which supports application privacy and code reuse. The Python translator and the extensive standard library are accessible in source or binary frame free of charge for every real stage and can be openly circulated (Python, 2001). Python is a universally useful programming language that can be utilised on any advanced PC working framework. It can be employed for handling content, numbers, pictures, logical information and saved content on a computer. Python is used on a day-to-day basis in the operations of the Google web crawler, the video-sharing site YouTube and also, Python assumes active parts in the achievement of business, government and non-benefit associations (Lukaszewski, 2017). A Python language is a good option in the field of big data analytics due to its simple usage and broad set of data processing libraries (Shaikh, 2016).

References

Adhikari, S., 2017. *9 REASONS WHY TABLEAU ROCKS.* [Online]
Available at: http://www.affecto.com/insights/blog/9-reasons-why-tableau-rocks/
[Accessed 28 April 2017].

Barbero, M. et al., 2016. *Big data analytics for policy making,* United Kingdom: A study prepared for the European Commission by Deloitte.

Batarseh, F. A., Yang, R. & Deng, L., 2017. A comprehensive model for management and validation of federal big data analytical systems. *Big Data Analytics,* 2(2), pp. 2-22.

Beal, V., 2016. *Big Data Analytics.* [Online]
Available at: http://www.webopedia.com/TERM/B/big_data_analytics.html
[Accessed 24 April 2017].

Bendre, M. et al., 2015. DATASPREAD: Unifying Databases and Spreadsheets. *The University of Illinois at Urbana-Champaign (UIUC),* 8(12), pp. 1-4.

Butler, 2016. *What is Tableau?.* [Online]
Available at: http://www.butleranalytics.com/what-is-tableau/
[Accessed 28 April 2017].

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

Chen, P. & Zhang, C.-Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Science,* 27(4), pp. 314-347.

Conyngham, G., 2017. *Lecture Presentation,* Dublin: Dublin City University.

Darden, M., 2016. *Five types of mobile analytics data and their uses.* [Online]
Available at: http://mobilebusinessinsights.com/2016/06/five-types-of-mobile-analytics-data-and-their-uses/
[Accessed 25 April 2017].

Deloitte., 2013. *The Analytics Advantage,* United Kingdom: Deloitte Touche Tohmatsu Limited.

EY & Nimbus Ninety, 2015. *Becoming an analytics-driven organisation to create value,* United Kingdom: Ernst & Young Global Limited.

EY, 2013. *Business opportunities: Big Data,* United Kingdom: Ernst & Young Global Limited.

EY, 2014. *Big data: Changing the way businesses compete and operate,* United Kingdom: Ernst & Young Global Limited.

Forfas, 2014. *Assessing the Demand for Big Data and Analytics Skills, 2013-2020,* Dublin: Skills Ireland.

Gartner., 2017. *Text Analytics.* [Online]
Available at: http://www.gartner.com/it-glossary/text-analytics
[Accessed 25 April 2017].

IBM, 2012. *What is big data?.* [Online]
Available at: https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
[Accessed 24 April 2017].

IBM, 2017. *What is Big Data Analytics?.* [Online]
Available at: https://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html
[Accessed 24 April 2017].

Interworks, 2017. *Data Visualisation.* [Online]
Available at: https://www.interworks.com/solutions/business-intelligence/visualization
[Accessed 28 April 2017].

Khanna, A., 2016. *Why Cloud Analytics is a Better Analytics.* [Online]
Available at: http://data-informed.com/why-cloud-analytics-is-better-analytics/
[Accessed 25 April 2017].

Lukaszewski, A., 2017. *What Is Python?.* [Online]
Available at: https://www.thoughtco.com/what-is-python-2813564
[Accessed 30 April 2017].

MacInnes, B., 2016. *Help SMEs benefit from big data on a small scale.* [Online]
Available at: http://www.computerweekly.com/microscope/feature/Help-SMEs-benefit-from-big-data-on-a-small-scale
[Accessed 26 April 2017].

Management Study Guide, 2017. *Database Concepts and its Application in HRIS.* [Online]
Available at: https://www.managementstudyguide.com/database-concepts-and-its-application-in-hris.htm
[Accessed 25 April 2017].

Manyika, J. et al., 2011. *Big data: The next frontier for innovation, competition, and productivity,* Newyork: McKinsey Global Institute.

Manyika, J. et al., 2011. *Big data: The next frontier for innovation, competition, and productivity,* California: McKinsey Global Institute.

OLAP.com, 2017. *What is the definition of OLAP?.* [Online]
Available at: http://olap.com/olap-definition/
[Accessed 25 April 2017].

Phippen, A., Sheppard, L. & Furnell, S., 2004. A practical evaluation of Web analytics. *Internet Research,* 4(284-293), p. 14.

Pointer, I., 2016. *Which freaking big data programming language should I use?.* [Online]
Available at: http://www.infoworld.com/article/3049672/application-development/which-freaking-big-data-programming-language-should-i-use.html
[Accessed 28 April 2017].

Python, 2001. *What is Python? Executive Summary.* [Online]
Available at: https://www.python.org/doc/essays/blurb/
[Accessed 28 April 2017].

Russom, P., 2011. *Big Data Analytics,* s.l.: The Data Warehouse Institute.

SAS & e-skills UK, 2013. *Big Data Analytics: Adoption and Employment Trends, 2012-2017,* United Kingdom: SAS: The Power to Know.

SAS, 2017. *Big Data Analytics: What it is and why it matters.* [Online]
Available at: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html
[Accessed 26 April 2017].

SAS, 2017. *Hadoop: What is it and why does it matter?.* [Online]
Available at: https://www.sas.com/en_ie/insights/big-data/hadoop.html
[Accessed 25 April 2017].

Schaefer, P., 2016. *What's the Difference Between Structured and Unstructured Data?.* [Online]
Available at: https://www.trifacta.com/blog/structured-unstructured-data/
[Accessed 25 April 2017].

Sena, D., Ozturk, M. & Vayvay, O., 2016. An Overview of Big Data for Growth in SMEs. *Science Direct: Procedia*, 23(5), p. 159 – 167.

Shaikh, S., 2016. *Why Python is important for big data and analytics applications?.* [Online]
Available at: https://www.eduonix.com/blog/bigdata-and-hadoop/python-important-big-data-analytics-applications/
[Accessed 28 April 2017].

Wagner, D., 2014. *The importance of big data analytics in business.* [Online]
Available at: http://www.techradar.com/news/world-of-tech/the-importance-of-big-data-analytics-in-business-1267606
[Accessed 20 April 2017].

Wall, M., 2014. *Big Data: Are you ready for blast-off?.* [Online]
Available at: http://www.bbc.com/news/business-26383058
[Accessed 24 April 2017].