www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-2, March-April 2017

International Journal of Engineering and Management Research

Page Number: 162-165

Deployment of Big Data and Analytics

Basanna Patagundi[1], Aparna Shenoy[2], Uday Tate[3]

[1,2]Associate Professor, School of Management, CMR University, Bangalore, INDIA

[3]Professor of Marketing, Lewis College of Business, Marshall University, Huntington, WV, USA

ABSTRACT

Big data and Analytics are new buzzwords in the organization. These are considered to be critical IT elements to sustain growth in the present business environment. However, organizations are struggling to adopt big data and analytics. Considering the availability of numerous big data and analytics tools available, the organization must have a broad operational model to integrate these tools with the existing system which would help them to reap the benefits of big data and analytics. This concept paper provides fundamental understanding of big data and analytics & its challenges and proposes a macro operational model for implementing big data in organizations.

Keywords-- Big data, Analytics, deployment model, operational model, big data capture and transmission, big data storage

I. INTRODUCTION

Big data and Analytics are evolving as two critical IT elements to sustain the growth in the present business environment. Big data is the process of capturing and storing useful business data and analytics refers to process of creating insights out of the data. Big data captures data from internal operations and external sources. The source of data is not just restricted to the internal transactions systems but also from external sources. With the advent of technological development such as Internet of Things (IoT), smart hardware, data capture at point of sales (POS) etc. the data from external sources can be captured relatively easily.

The average companies have 427 times the amount of data ever recorded in the Library of Congress. Facebook has more photographic data than all pixels ever processed by Kodak. As per International Data Corporation (IDC) data from 2013 to 2020, the digital universe will grow by a factor of 10 – from 4.4 trillion gigabytes to 44 trillion. It is growing more than double every two years.

Analytics gained momentum in the recent past. Thomas Davenport classifies analytics phases in two eras namely, before big data(BBD) and after big data (ABD). According to Davenport, there are three phases of Analytics namely, Analytics 1.0, Analytics 2.0 and Analytics 3.0. Business intelligence constituted analytics 1.0.The data about operations within the organization was captured, aggregated and analyzed to make informed decisions by Managers. Business intelligence generated reports and aggregated data about what happened in the past. Analytics 2.0 was driven by e-commerce and social media. These sites generated huge amount of data about customers' transactions and all other activities generated across the value chain. The huge amount of data was named as Big Data. The technology of Hadoop helped to store huge amount of data generated in the organization. The traditional database management system did not have the capability to support the massive amount of unstructured data. Hadoop has the capability of storing huge data which is either structured or unstructured. Analytics 3.0 was the era of data-enriched offerings. "The pioneering big data firms in Silicon Valley began investing in analytics to support customer-facing products, services, and features. They attracted viewers to their websites through better search algorithms, recommendations from friends and colleagues, suggestions for products to buy, and highly targeted ads, all driven by analytics rooted in enormous amounts of data." (Davenport, 2013). The data is captured in every business transactions enabling all firms to provide better offerings to the customers. Data science is a methodology that helps extract knowledge from data. BI tools such as Microsoft SQL Server Analysis and reporting services, SAP Business Objects, IBM Cognos/SPSS, SAS and Micro strategy use techniques such as data mining, cluster analysis and predictive modeling to analyze the transactional data. Leading social

listening tools such as Big Data InfoSphere, IBM Watson and Oracle Endeca use techniques such as crowdsourcing, textual analysis, sentiment analysis and network analysis to analyze the non-transactional and social media data.

II. BIG DATA AND ANALYTICS IN PRACTICE

In one of the posts on social media Bernard Marr one of the influential thinker in Big Data mentions that worldwide, companies are expected to spend £80 billion on products and services to enable them to carry out big data analysis this year. But when it has been estimated that failing to effectively capitalise on their data costs business in the US alone close to £400 billion each year, it's no surprise that respected market researchers Gartner recently found that 73% of businesses have already invested in a big data plan, or are planning to do so in the next few years.

The only way to remain competitive is to capture, manage and use the data generated. "Most business owners are often overwhelmed by the seemingly Herculean task they face of extracting meaningful information from the sea of data. They read stories about the eye popping results that other organizations have enjoyed and how they never discard any data, how everything in these data rich companies is captured and analyzed to within an inch of its life because it's valuable and potentially offers unique and powerful insights for business development." (Marr, 2015) Canada based consulting and system integration company, CGI has integrated IBM's applied Customer Insight portfolio which has enabled CGI to capture data acquired from multiple customers' touch points, and interpret data to get more customer insights and implement it collaboratively across the whole business operation. Understanding the behavior and sentiment of customers requires social listening, that is, capturing what is said about an organization in the social world. Big data is the fire that ignites IBM's Applied Customer Insight portfolio that would help understand the behavior and sentiment of customers.

Idea Cellular uses big data analytics to help acquire profitable customers, retain them, up-sell and cross-sell, and engage them in meaningful interactions. Today the company harnesses a massive amount of data, which is generated from its customer base of more than 150 million subscribers serviced by more than 110,000 telecom towers. Idea Cellular has nearly 1 billion transactions per day, which results in 65TB of data per month. Idea Cellular uses the SAS Marketing Automation Framework for data management and campaign management. This deals with huge volume of data to configure business rules and seamless integration with other operational systems.

HDFC Bank has adapted sophisticated information technology to become a world-class provider of wholesale and retail financial services. SAS, combined with the bank's CRM solution, helps HDFC Bank to model its customer data by assigning propensity to buy, spend and activate which helps the bank target sales communications to its customers thereby reducing the number of calls each customer receives.

The implementation of big data and analytics is creating value to the organization. However, the organizations are facing a few of the challenges as mentioned below.

- Measuring the value created by big data and analytics
- Lack of clarity in the deployment model of big data and analytics

III. IMPLEMENTATION OF BIG DATA AND ANALYTICS

Companies are turning to analytics solutions to extract meaning from the huge volumes of data to help improve decision making. The current literature has identified three broad types of Analytics as follows.

3.1 Descriptive Analytics: This uses data aggregation and data mining techniques to provide insight into the past and answer: "What has happened?"

3.2 Predictive Analytics: This uses statistical models and forecasts techniques to understand the future and answer: "What could happen?"

3.3 Prescriptive Analytics: This uses optimization and simulation algorithms to advice on possible outcomes and answer: "What should we do?

There are many software tools available for analytics, the predominant tools in analytics are SAS, R-tool and IBM – Cognos BI, SPSS and Biginsight infosphere.

Organizations today struggle to get started with Analytics. There is a concrete realization that analytics is necessary for business success. Yet, a lot of organizations face material challenges while starting their analytics journey.(Kruxonomy) Bernard offers the SMART model offers a much needed way to navigate the oceans of data to find the pockets of meaning that can transform productivity, efficiency, performance and, ultimately, bottom line results. S = Start with Strategy, M = Measure Metrics and Data, A = Apply Analytics, R = Report Results, T = Transform your Business. Many approaches similar to SMART model are mostly at the strategic level. However, there is a need of a generic operational model to implement big data in the organization.

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

IV. PROPOSED MODEL

Below figure broadly describes a broad operational model that can be used to capture and make of big data.

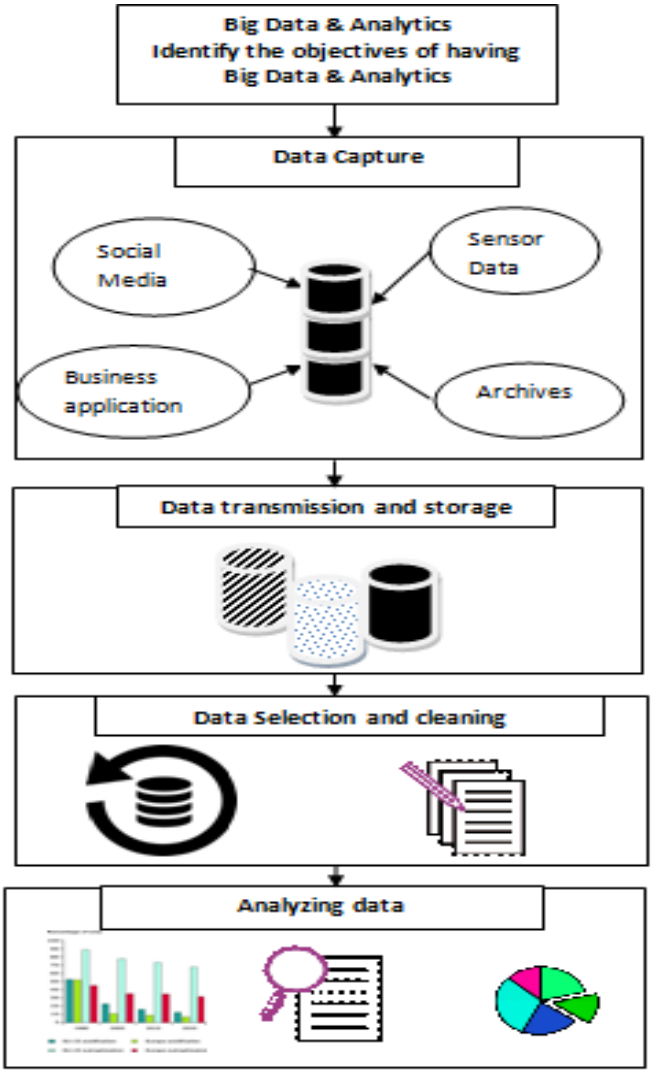

Fig:4.1 Macro view of operational model for Big Data and Analytics

4.1 Big Data & Analytics: Identify the objectives of having Big Data & Analytics

The organization has to identify the purposes of having big data & analytics. This provides the appropriate directions to further plan for identifying the required data and sources of data to be collected. This phase gives basic direction to the organizations to navigate through the big data and analytics.

4.2 Data Capture:

The data source module captures the structured and unstructured data from various sources at varying velocity. Business transactions are the key sources of data. Transactional data consists of structured, low-level, detailed information capturing key characteristics of a customer transaction such as purchase, claim, cash transfer, and credit card payment. Unstructured data can be in the form of documents such as emails, web pages and feedback forms. The organisation has to create a checklist of the touch points of stakeholders where data is generated. The data generated across these touch points have to be captured and transmitted for analysis.

4.2 Data transmission and storage:

Data is split or combined on the basis of its type, contents and the requirements of the organization. Hadoop is an open source framework used to store large volumes of data in a distributed manner across multiple machines. A scalable Hadoop Distributed File System (HDFS) supports petabytes of data. The growing volume of data

can be difficult to store and handle. HDFS can be implemented in an organization at comparatively low costs which can handle continuous streaming of data.

4.3 Data Selection and Cleaning:

The data quality is very important for further analysis. High- quality data in combination with good technology results in better analysis and insights. Data cleaning is required when integrating heterogeneous data sources. The information relevant for the organization is filtered on the basis of master data repository. The data is cleansed by removing the noise and minimizing redundancy.

4.4 Analyzing data:

The role of this module is to analyze large amount of structured, semi-structured and unstructured data. The analysis is related to text analytics and statistical analytics. Following are the various methods used for analysing big data:

- Natural language processing
- Text mining
- Machine learning
- Search and sort algorithms

Map Reduce simplifies the creation of processes for analyzing huge amounts of structured and unstructured data. Map is a component that distributes a problem across a number of systems and also acts as a recovery management system in case of failure. When the task of distribution computation is completed, the reduce function combines all the elements back together to provide aggregate result.

4.4.1 Deriving Conclusions:

Customized reports can be generated including comparative analysis using current and historical data. Analysis of transactional data provides organizations with actionable insights into their operations, helping them improve their working capital management, increase customer acquisition, retention, and satisfaction, among many other things.

V. CONCLUSION

Big data and analytics are critical components in the organization. The data is growing at a faster rate than anticipated. Therefore, the organizations must be equipped with the processes and tools to assimilate huge data which is referred as "Big Data". This paper has developed a broad operational model which can be used to implement big data and analytics in the organisations. The model has four broad processes namely, data capture, data selection and cleaning and analyzing of data. The organization can make use of this broad model to implement big data and analytics successfully.

REFERENCES

[1] Barabási, A. 2003.,Linked: *How Everything Is Connected to Everything Else and What it Means for Business*, Science, and Everyday Life, New York: Plume

[2] Bloomberg Businessweek. 2011. The Current State of Business Analytics: Where Do We Go from Here?, Bloomberg Business- Chen et al./Introduction: Business Intelligence Researchweek Research Services (http://www.sas.com/resources/asset/busanalyticsstudy_wp_08232011.pdf)

[3] Chaudhuri, S., Dayal, U., and Narasayya, V. 2011. *An Overview of Business Intelligence Technology* Communications of the ACM (54:8), pp. 88-98.

[4] Chen, H. 2011a. *Design Science, Grand Challenges, and Societal Impacts* ACM Transactions on Management Information Systems (2:1), pp. 1:1-1:10

[5] Davenport, T. H. 2006. *Competing on Analytics*, Harvard Business Review (84:1), p. 98-107

[6] Davenport Thomas H.. 2014. *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities.* Harvard Business Review Press

[7] Hevner, A., March, S. T., Park, J., and Ram. S. 2004. *Design Science Research in Information Systems* MIS Quarterly (28:1), pp. 75-105.

[8] Henschen, D. 2011. *Why All the Hadoopla*? Information Week, November 14, pp. 19-26

[9] March, S. T., and Storey, V. C. 2008. "Design Science in the Information Systems Discipline," MIS Quarterly (32:4), pp. 725-730

[10] Hsinchun Chen , Roger H. L. Chiang , Veda C. Storey, Business intelligence and analytics: from big data to big impact, MIS Quarterly, v.36 n.4, p.1165-1188, December 2012

[11] Marr Bernard. 2015.*The Bernard Marr column: Why big data matters to every business.* Retrieved from www.Hiscox.co.uk on June 20, 2015.

[12] Marr Bernard . 2015. *Take a SMART Approach to Big Data Analytics.* Retrieved from http://data-informed.com on June 20,2015.

[13] March, S. T., and Storey, V. C. 2008. *Design Science in the Information Systems Discipline.* MIS Quarterly (32:4), pp. 725-730

[14] Russom, P. 2011. *Big Data Analytics*, TDWI Best Practices Report, Fourth Quarter.

[15] Watson, H. J., and Wixom, B. H, 2007, *The Current State of Business Intelligence* IEEE Computer (40:9), pp. 96-99.

[16] Parise Salvatore , Iyer Bala , Vesset Dan , *Four strategies to capture and create value from big data,* Ivey Business Journal, August 2012.