IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 5, Ver. III (Sep. - Oct. 2016), PP 61-67
www.iosrjournals.org

Big Data: A Stimulus in Business Analytics

Amita Garg[1], Dr. Mamta Padole[2]
[1]*Parul Institute of Management & Research, Parul University, India*
[2]*Faculty of Technology and Engineering, The Maharaja Sayajirao University of Baroda, India*

Abstract: *Harvesting data and investing in analytics have almost become a norm of the organization. Successful organizations are achieving business advantages by analysing massive amount of data that is generated during their business practices. This unconventional enormous size data is referred as Big Data. Big data has received significant attention in recent years. Although, Big Data Analytics has become the need of the day, several businesses are still lagging. The main issues that limit the organizations to begin using big data analysis are Lack of knowledge to handle challenges of big data, Lack of knowledge about advantages of big data and Non-Awareness of its impact on business. In this paper, an attempt is made to study the importance of big data analytics in the business.*

Keywords: *Big data, Big Data Analytics, Business Analytics, Business Intelligence*

I. Introduction

The aim of big data analysis is to help in making the process of decision making efficient. Efficiency means providing results that are accurate, timely and has future insight. "Big Data" indicates data that is huge. Daily, the world creates 2.5 quintillion bytes of data, 90% of the data in the world today has been created in the last two years itself [1]. Data is growing at an exponential rate and the experts of the data analytics technology are still working on to acquire enough knowledge to analyse this enormous amount of data. Organizations are motivated to go for big data analysis so that it enables the organizations to help in following:

1. To organize the huge data.
2. Analysis of organized data provides new business opportunities.
3. Take advantages of existing technologies that support business data analysis.

Big data comprises of structured, unstructured and semi structured data. The data content may be of type Text, Audio, Video and Image. The paper documents the basic concepts, metrics and characteristics of big data. It also emphasises on analytic techniques on these various data types.

This study is an attempt to investigate the role & goal of big data analysis in business, types of big data analysis techniques for gaining business intelligence, key challenges faced by businesses in using business data analysis and solutions thereof.

II. Big Data

The novel concept of big data is derived from the world of computer science and econometrics [2]. Mc Kinsey & Company is the first company to refer to the term big data. Since then big data draws concerns of various types of industries. There are many definitions of big data. Big data includes structured and unstructured data sets having huge size which is beyond the ability of frequently used software tools to gather, curate, manage, and process in a specific time [3]. International Data Corporation (IDC) is defined as to meet 4V's (Variety, Velocity, Volume, Value) index called big data. The characteristics of big data are submitted by Victor Meyer-Schonberg in "BIG DATA" [4]. (These four characteristics are: volume, velocity, variety and value) Volume refers to the huge amount of data. With the advancement in data storage and network technology, data storage expands from TB to ZB. In 2011 itself, 1.8ZB (1.8 trillion GB) of data was created. Warehouse management servers have increased from 10-fold to 50-fold to cater the speedy growth of big data. Velocity refers to the mobility of data streams. It is difficult to deal with data in a traditional way because data needs to be processed in real time. Variety refers to relational and non-relational data generated by different means. With the development of mobile networks, people are more prone to use real-time data. The quantity of semi-structured and unstructured data is also increasing. Value reflects the significance of big data applications, which process scarce, uncertain and diverse data. Big data analysis focuses on data collection, storage & analysis to extract out value from data. Big data emphasizes on complexity in data analysis, and it pays more attention to data processing efficiency and the value of data.

III. Big Data Analytics

Big data are useless until its potential value is reflected in using it in business decision making. To gain BI, organizations need efficient processes to turn high volumes of fast-moving and diverse data into meaningful insights. The whole process of extracting insight from big data can be implemented in five stages (Labrinidis & Jagadish, 2012), as shown in Fig. 1. These five stages may be divided as two main sub-processes: data management and data analytics. Data management comprises of processes and supporting technologies to acquire and store data and pre-processing this data for analysis. Analytics, on the other hand, refers to techniques used to analyse and acquire business intelligence from big data.

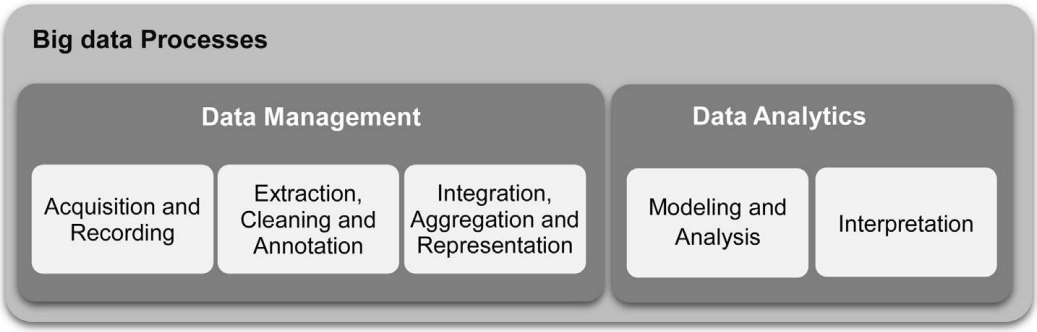

Figure 1. Five Stages of Big Data Analysis **[**Labrinidis & Jagadish, 2012]

3.2 Text Analytics:

Text analytics is technique that extracts information from textual data. Text analytics enable businesses to convert large volumes of text generated by human text into meaningful summaries, which support evidence-based decision-making. Some of the text analytics methods are:

a. Information Extraction (IE) techniques: It is used to extract structured data from unstructured text. IE consists of two sub-tasks Entity Recognition (ER) and Relation Extraction (RE)[5] .
b. Text summarization techniques: It automatically produces a brief summary of a single or multiple documents summarization using two approaches: the extractive approach and the abstractive approach. In extractive summarization, summary which is a subset of original text is created. In contrast, abstractive summarization techniques involve extracting semantic information from the text. In order to parse the original text and generate the summary, it incorporates advanced Natural Language Processing (NLP) techniques.
c. Question Answer (QA) techniques: It provide answers to questions posed in natural language. Like abstractive summarization, QA systems rely on complex NLP techniques. QA techniques are further classified into three categories:
i. Information retrieval (IR) based approach
ii. Knowledge-based approach
iii. Hybrid approach.

The Knowledge-based QA systems are particularly helpful in restricted domains, such as tourism, medicine, and transportation, where large number of documents do not exist. Such domains lack data redundancy, which is required for IR-based QA systems.

d. Sentiment analysis (opinion mining) techniques: It analyzes opinionated text containing people's opinions about objects such as products, organizations, individuals, and events. Marketing, finance, and the political and social sciences are the major areas of applications for sentiment analysis. Sentiment analysis techniques are further divided into three sub-categories:
i. Document-level
ii. Sentence-level
iii. Aspect-based.

3.2 Audio Analytics

Analysing and mining information from audio data is referred as audio analytics. As it is mostly been applied to spoken audio, the terms audio analytics and speech analytics are often used interchangeably. Speech analytics uses two approaches: the transcript-based approach (large-vocabulary continuous speech recognition, LVCSR) and the phonetic-based approach.

a. LVCSR systems use a two-phase process: indexing and searching. In the first phase, the speech contents of the audio are transcribed by using automatic speech recognition (ASR) algorithms. This identification of

words is done based on a predefined dictionary. If the system fails to find the exact word in the dictionary, the most similar word is returned. This generates a search index file that contains information about the sequence of the words found in the speech. The second phase then uses a standard text-based method to find the search term in the index file.

b. A Phonetic-based system deals with sounds or phonemes. Phonemes are the distinct unit of sound in a specified language that discriminates one word from other. A Phonetic-based system also uses two phases: phonetic indexing and searching.

3.3 Video Analytics

Video analytics, also known as video content analysis (VCA), is a technique to monitor, analyze, and extract meaningful information from video streams. One second of a high-definition video, is equivalent to over 2000 pages of text in size[6] . Also YouTube Statistics and recent studies reveals that 100 hours of video are uploaded to YouTube every minute. Big data technologies turn this challenge into opportunity by automatically examining through and drawing intelligence from thousands of hours of video. There are two methods of video analytics: server-based and edge-based:

a. Server-based architecture. Here, the captured videos are routed back to a centralized and dedicated server that performs the video analytics. The limitation in this approach is that due to bandwidth restrictions, the video needs to be compressed by reducing the frame rates and/or the image resolution. The resulting loss of information can affect the accuracy of the analysis.

b. Edge-based architecture. Here analytics are applied on the 'edge' of the system means analytics is performed on the raw data captured by the camera. This method is more effective for analysis as the entire content of the video stream is available here. However, this is more expensive to maintain.

3.2 Image Analytics

Image analysis is the extraction of meaningful information digital images by means of digital image processing techniques. The use of bar codes and QR codes are simple examples of Image Analytics.

To a computer, images are either a raster image or a vector image. Raster images are a sequence of pixels with discreet numerical values for colour. Vector images are a set of colour-annotated polygons. Analytics on images or videos are performed by the geometric encoding that transformed into constructs depicting physical features of objects and movement represented by the image or video. These constructs can then be logically analyzed by a computer.

Image analytics system in transforming images and videos to analytically prepare a dataset (a set of time series, one per variable) follows the steps as given in fig 2.[7]

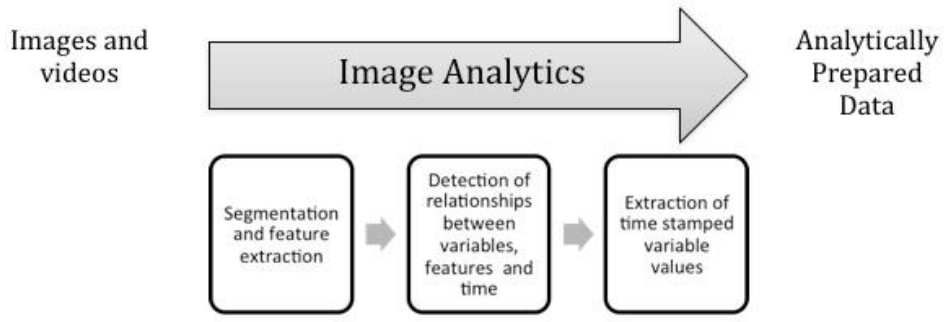

Figure 2. The role of Image Analytics in Transforming Images and Videos to analytically prepared data.

The first transformation step segments images into structured elements and prepares them for feature extraction – i.e., the identification of low-level features in the image. The second transformation step is the detection of relationships between these features, variables and time. The third transformation step is the extraction of variables with time-stamped values.

Examples of image analysis techniques in different fields include:

- 2D and 3D object recognition,
- motion detection e.g. Single particle tracking,
- optical flow
- medical scan analysis,

IV. Big Data Analytics & Business Goals

Big data analytics is helping business organizations in achieving the following goals:

1. Impacting customer relationships: Big data is helpful in developing micro customer segmentation, proximity based marketing, real-time relevant offers, high tendency cross-sell recommendations and sentiment analysis. By cross-analysing store and online interactions and conversions, and further cross-referencing the results by consumer demographical and geographical data, retailers will discover with far greater accuracy how to pinpoint the ideal customers for select products, deliver messaging for improved engagement and create offers for improved conversions.
2. Redefining product development: Information on consumer preferences and manifestations in communities and forums are inputs for the development and adjustment of products, as well as for the definition of complementary services. Product development on the base of customer's ideas and evaluation in real time is possible.
3. Reorganizing business operations: In many manufacturing companies, IT systems are not integrated, which makes it difficult for cross-functional teams to accomplish shared goals. Big data analytics helps in gaining intelligence about business operations and also optimizes production workflows, inventory, Work In Progress (WIP), and value chain decisions.
4. Shifting in focus from experiential driven business to data-driven business: Big data shifts the focus of various businesses. Earlier all types of business decisions were done either by experience or by use of limited data. But due to advancement in various big data analytics techniques decisions are taken based on data available from various source and in large quantity.
5. Optimizing business through supply chain management: Applying big data to demand planning can aid just-in-time inventory distribution and improved logistics to help get the right products to the right destinations at the right time – and reduce both overstocks and stock outs.

V. Big Data & Business Case Studies

Organizations are trying to understand concepts of big data, its impact & benefits in business organizations. A survey reveals that the only 12 percent organizations are implementing or executing the big data strategy and 71 percent organizations are in the planning stage [2]. Big data analytics are helping organizations to gain Business Intelligence (BI) in improvising business strategies to make profit & to compete better with other organizations. Types of decisions that organizations can make using big data analytics are smarter, future oriented and decisions to gain competitive advantage.

There are four types of big data BI that really help businesses:

1. Prescriptive – This type of analysis reveals what actions should be taken. This is most important analysis and usually results in rules and recommendations for next steps.
2. Predictive – This analysis helps to identify the commonly occurring scenarios that might also occur in future. The deliverables are usually a predictive forecast.
3. Diagnostic – This approach uses the past data to determine what happened and why. The result of the diagnostic analysis is helpful in identifying past errors, if any or analysing the past approach.
4. Descriptive – This analysis helps of know what is happening currently based on incoming data. This describes the current scenario of the business.

Applications: Big data analysis can be applied in the following business sectors:

1. **Financial Sector:**

- Big data analytics helps in predicting how customers and competitor's customers will behave and how that behaviour will help in critically tailoring and pricing products.
- Data analytics in banking helps in recovery of bad debt. Recoveries functions generally are based on the delinquency status of the account. However, a better understanding about customer circumstances can improve targeting and impacts on recovery rate.
- The Securities Exchange Commission (SEC) is using big data to monitor financial market activity. They are currently using network analytics and natural language processors to catch illegal trading activities in the financial markets.
- This industry also heavily relies on big data for risk analytics including anti-money laundering, fraud mitigation and in identifying their customers through (KYC) Know Your Customer.

2. **Healthcare Sector:**

- Big data analytics in health care helps to reduce cost of treatment, predict outbreaks of epidemics, avoid preventable diseases and improve the quality of life.
- Big data analytics in health care contributes in Genomic analytics. Genomic analytics execute gene sequencing more efficiently and cost effectively and make genomic analysis a part of the regular medical care decision process.
- Patient profile analytics: This is applied to patient profiles (e.g., segmentation and predictive modelling) to identify individuals who would benefit from proactive care or lifestyle changes, for

example, those patients at risk of developing a specific disease (e.g., diabetes) would benefit from preventive care.

3. Agricultural sector:

- Big data Farming, also called 'precision farming' is expected to play an important role by weather forecasting, irrigation scheduling, and optimization of farming machinery and monitoring of grain prices.
- Mining big data of existing crop, soil and climate data and analysing data help in optimizing production and make agriculture more resilient to climate change

4. **Energy & Natural resources:**

- Exchange-rate margin analysis helps to determine where to buy the raw material and how to price the products.
- Big data analytics is helping this business sector in better maintenance management. According to Global Data a U.K. consulting firm capital expenditure in Oil & Gas business is increasing every year. To slow down this increasing growth rate companies have started using predictive analytics to determine when to go for cleaning of device or replacement. E.g., Qatargas, World's largest liquefied natural gas producer is using predictive analysis with in-memory computing for cutting operational maintenance cost. [8]

5. **Communication Media & Entertainment**: [9]

- Predicting what audiences want: The scope of big data collected by the media and entertainment industry and the potential to mine it to understand what content, shows, movies and music consumers want to view and hear most.
- Scheduling Optimization: Using insights from big data, media and entertainment companies are able to understand when customers are most likely to view content and what device they will be using when they view it. With big data's scalability, this information can be analyzed at a granular ZIP code level for localized distribution
- Increasing Acquisition and retention: By using big data to understand why consumers subscribe and unsubscribe, media and entertainment companies can develop the best promotional and product strategies to attract and retain customers.

Case Studies:

1. UPS (United Parcel Service)

- UPS is capturing big data by tracking variety of parcel movements. This big data is coming from telematics sensors used with over 46,000 vehicles.
- The data on UPS package cars (trucks), for example, includes their speed, direction and braking. The data is not only used to monitor daily performance, but to drive a major redesign of UPS drivers' route structures.
- ORION (On-Road Integrated Optimization and Navigation), the world's largest operations research project initiated by UPS relies heavily on online map data, and will eventually reconfigure a driver's pickups and drop-offs in real time.
- The project has already led to savings in 2011 of more than 8.4 million gallons of fuel by cutting 85 million miles off of daily routes. UPS estimates that saving only one daily mile driven per driver saves the company $30 million, so the overall dollar savings are substantial. The company is also attempting to use data and analytics to optimize the efficiency of its 2000 aircraft flights per day. [10]

2. Caesars Entertainment Corporation:

- It is an American Gaming corporation who adopted data analytics, particularly in the area of customer loyalty, marketing, and service.
- Caesars uses big data technologies to respond in real time for customer marketing and service. For example, the company has data about its customers from its Total Rewards loyalty program, web clickstreams, and from real-time play in slot machines. It has traditionally used all those data sources to understand customers, but it has been difficult to integrate and act on them in real time, while the customer is still playing at a slot machine or in the resort. In order to pursue this objective, Caesars has acquired both Hadoop clusters and open-source and commercial analytics software.
- With video analytics on big data, Caesars employ more automated means for spotting service issues involving less frequent customers.
- Caesars also begin to analyze mobile data, and is targeting real-time offers to mobile devices. [10]

3. European Mobile Telecom Group:

- It is among top five telecom companies worldwide. The various challenges they are facing are to track performance and active users of mobile data services. Company also want to capture data on mobile usage and customer activities.

- Big data analysis of service usage would enable the company to spot upcoming trends and intelligently market them to customers, and keep its customer touch points current. Also, being better informed about customer usage would enable the enterprise to ensure it supported all of the latest mobile phone technologies seen on its network and be alerted to potential incompatibilities.
- For this company start using Hadoop for big data access and analytics which enables the business to report on usage, and provide other KPI information. Hadoop allows huge amounts of data to be stored in a granular fashion that is cost effective and performant. [10]

4. Schneider National:

- It is one of America's largest truckload, logistics and intermodal services providers.
- Schneider started using low-cost sensors for its trucks, trailers and intermodal containers. The sensors monitor location, driving behaviours, fuel levels and whether a trailer/container is loaded or empty.
- Schneider transition to big data analytics help in optimizing decisions it makes with the sensor data. For example dispatching of trucks and containers is improved substantially, and the company's use of prescriptive analytics is changing job roles and relationships.
- Safety is a core value at Schneider. Driving sensors are triggering safety discussions between drivers and their leaders. Hard braking in a truck, for example, is captured by sensors and relayed to headquarters. This data is tracked in dashboard-based safety metrics and initiates a review between the driver and his/her leader. [10]

VI. Big Data Challenges

1. Variety of Data Management [11]
   Wide varieties of data management tools supporting both operational and analytical processing are available in market. Each tool is suited differently, some tools provide greater flexibility; others are eminently scalable in terms of performance while others support a wider range of functionality. So Business organizations are facing uncertainty in selecting right tool.
2. Availability of big data skill set
   Business organizations are motivated for big data applications for which broad community of experts are required. Unavailability of experts having required skill set creates a challenge.
3. Convergence of variety of data
   The scale and variety of data to be used into a big data environment is very large. This imposes a convergence problem and making data accessibility and integration our third challenge.
4. Synchronization of Data Sources
   Data sets from diverse sources are combined into an analytical platform. These data sets inconsistencies and asynchrony pose a risk of faulty analytical results and have a much disastrous effect.
5. Seamlessness & Transparency
   To enable data accessible to the different downstream applications in seamless and transparent way, pose a challenge.

VII. Conclusion

In this research, an innovative topic of big data is examined, which has recently gained lots of interest due to several opportunities and benefits. In the information era we are currently living in, voluminous varieties of high velocity data, and within them laid basic details and patterns of hidden knowledge which is of great significance to business organizations. Hence, big data analytics can be applied to influence businesses and enhance decision making. Big data, its characteristics and importance and different big data analytics techniques are examined. By applying these analytics to big data, valuable information can be extracted and exploited to enhance decision making and support informed decisions. Consequently, some of the different areas where big data analytics can support and aid in decision making were examined. It was found that big data analytics can provide vast horizons of opportunities in various applications and areas, such as customer intelligence, fraud detection, and supply chain management. Additionally, its benefits can serve different sectors and industries, such as healthcare, retail, telecom, manufacturing, etc. This gives various business organizations an idea about importance of big data analysis. It aids in developing an idea of what can be done to provide more enhanced solutions for big data analytics that support in decision making. Finally, any new technology, if applied correctly can bring with it several potential benefits as well as challenges. Various challenges that emerged are require high skill set, techniques for synchronization, data transparency and integration. Therefore, future research can focus on providing a roadmap or framework for big data management which can help to overcome the mentioned challenges

References

[1]. J. P. Dijcks, *Oracle: Big data for the enterprise*, *Oracle White Paper*, 2012.
[2]. Big Data Survey Research Brief, *Sas White Paper*,2013.
[3]. M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report, *IBM Institute for Business Value and Said Business School at the University of Oxford*, 2012.
[4]. Turn Big Data into Big Value, A Practical Strategy, *Intel White Paper*, 2013.
[5]. JIANG, Jing. Information Extraction from Text. (2012). *Mining Text Data*. Research Collection School Of Information Systems.
[6]. Manyika, J., Chui,. M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, June). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
[7]. Analytics-magazine.org/images-a-videos-really-big-data
[8]. http://www.sap.com/bin/sapcom/en_us/downloadasset.2013-09-sep-25-10.big-returns-from-big-data-in-energy-and-process-industries-pdf.bypassReg.html
[9]. https://www.qubole.com/resources/solution/big-data-media-entertainment/
[10]. Tom Davenport and Jill Dyché and SAS Institute Inc., Big Data in Big Companies
[11]. David Loshin, President of Knowledge Integrity, Inc. *Addressing five emerging challenge*, www.progress.com