

# BIG DATA – INTRODUCTION, APPLICATIONS AND FUTURE SCOPE

Shruti Ahuja  
Assistant Professor  
D. A. V College, Abohar,  
Punjab, India

## ABSTRACT

The term Big Data refers to the huge amount of data that cannot be dealt with by traditional data-handling techniques. Big Data is still a unique concept, and this article is intended to elaborate it in depth. The study goes on to reveal the applications of Big Data in all different aspects of economy and being. The employment of Big Data Analytics after its integration with digital capabilities to safeguard business growth and its visualization to make it understandable to the technically apprenticed business analysts has been discoursed in detail. Aside this, the incorporation of Big Data in order to improve health of humans, for the benefit of finance, food industry, telecom industry and for fraud detection have been explained. Inability in finding sound data professionals and software that possess ability to process data at a high speed is also covered.

## INTRODUCTION

Every day, 2.5 quintillion bytes of data is created. This data comes from everywhere: from sensors used to collect climate information, digital pictures and videos, posts at social media sites, records of purchase transactions etc. [1]. Such enormous amount of data produced continuously is coined as Big Data. Big Data decodes previously safe and sound data to derive new visions that gets incorporated into business operations. However, as there is exponential increase in amounts of data, the current techniques are becoming out-dated. Dealing with Big Data requires extensive coding skills and domain knowledge. Despite being Colossal in nature, applications of Big Data are almost ubiquitous- from marketing applications to scientific research to customer interests and so on. Today, we can witness Big Data in action almost everywhere. e World Bank organized the first WBG Big Data Innovation Challenge which got several unique ideas applying Big Data in poverty prediction and for climate smart agriculture and in user focused Identification of Infrastructure Condition of roads and their safety and so on [2]. At present times, the colossal amounts of data are being generated at unbelievably high speeds. Taking Facebook [3] for example – it generates 2.7 billion actions of like per day and 300 million photos amongst others roughly amounting to 2.5 million sections of content in each day while Google Now processes over 1.2 trillion searches per year worldwide. [4]. Implementing Big Data is a massive task as it requires the large volume, velocity and variety. Big Data is a term covering the use of techniques to capture, process, analyse large datasets in a reasonable timeframe which is otherwise not accessible to standard IT technologies. The platform, tools and software used for this purpose are jointly called —Big Data technologies. [5] Currently, Hadoop is the most commonly implemented technology. Hadoop is the result of several other technologies like Hadoop Distribution File Systems, Pig, HBase and hive etc. However in near future, Hadoop and other existing methods will be incapable of dealing with the complexities of Big Data. At present, there are two general approaches to big data. They are:

*Divide and Conquer using Hadoop:* The massive data set is broken into smaller parts and those parts are processed in a parallel fashion using many servers.

*Brute Force* using technology on the likes of SAP HANA: This includes a powerful server with huge storage and is used to compress the given data set into a single unit.

## APPLICATIONS

Big Data have become ubiquitous. On the grounds of business, health or general living standards big data analytics are applicable. In simple terms, Big Data is a field which can be used in any area where large quantity of data is to be handled. The major applications of Big Data have been listed below:

### *Data Visualization*

Organizations throughout the world are perpetually getting familiar with the importance of big data analytics. From predicting the customer purchasing behaviour patterns to inducing them to make purchases to detecting fraud and misuse which was used to be an perplexing task for most companies big data analytics is a one-stop solution. Business experts need to have the opportunity of questioning and interpreting the data according to their business necessities irrespective of the volume and complexity of the data. In order to achieve this, data scientists need to visualize and present this data in an understandable way. Giants like Google, Twitter, eBay, Wal-Mart, Facebook, etc., adopted data visualization to ease up the complexity of handling data. Data visualization has shown enormous positive results in various business organizations. Implementing data analytics and data visualization, organisations can finally begin to tap into the immense potential of Big Data and ensure greater business stability and return on investments.

### *Integration*

As need of the 21st century integrating digital capabilities in decision-making is bringing transformation in the organizations. By transforming the processes, various companies are developing flexibility and precision that allows new growth. Using social and mobile technologies to modify the way people connect and interact with the organizations and integrating big data analytics in the aforesaid process is proving to be a boon for such organizations who are implementing it. With the use of this concept, enterprises have found ways to leverage the data better for increasing revenues or to cut the costs. Customer-centric objectives are still the primary concern of most businesses, giving a shift to integrating big data technologies into the background operations and internal processes.

### *Healthcare*

Healthcare is one of those fields in which Big Data have the maximum social impact. Right from the diagnosis of potential health hazards in the individuals to complex researches in medical science, big data is present in all aspects of it [6]. Devices such as the Jawbone [7] and the Samsung Gear Fit [8] allow the users to track and upload data. Such data will get compiled and made available to doctors, which aid them in the diagnosis. Several partnerships like the Pittsburgh Health Data Alliance have been established. Health care field generates an enormous amount of data every day. There is a need to mine this data and provide it to the medical practitioners and researchers who can put it to work in real life, to benefit people. The solutions we develop will be focused on preventing a disease, improving the diagnosis of such diseases and enhancing the quality of care provided. Further, there exists a possibility to lower health care costs, which is a greatest challenges faced by the people of our nation. The patient's diagnosis could be analysed and compared with the symptoms of other patients to discover the patterns and to ensure that they have been given the best treatment.

### *World of Finance*

Big Data can be a useful tool in analysing the extremely complex stock market moves and it also aids in making universal financial decisions. For example, an extensive analysis of the big data available on Google Trends can provide assistance in forecasting the stock market. Though this is not a sure method, it definitely

is an advancement in the field. Big Data has also being implemented in a field called Quantitative Investing [9] where data scientists with minor financial training are trying to incorporate computing power into predicting securities prices by getting ideas from various sources like newswires, earning reports, weather bulletins, Facebook and Twitter. An exciting opportunity of using Big Data in finance is the sentiment extraction [10] from news articles. Market sentiment refers to the irrational belief in investors about cash-flow returns [11].

#### *Fraud Detection*

Forensic Data Analytics or FDA has been an exciting area of interest and few companies are actually using FDA for big data mining. The reasons behind this situation vary from the lack of awareness and expertise, for developing right tools to mine big data. This is all because of lack of suitable technologies and inability to handle such enormous quantities of data. Corporations know that there are high risk numbers in book entries, such as thousands or duplicates, but they have just started analysing the descriptions for those book entries. Looking at numbers and words can mean the difference between recognition of fraud, and falling victim to it. The combination of suitable data and big data analytics can help combat fraudulent activities to some extent. Another universal approach for implementation of big data analytics is required and is developed by Companies such as Pactera [12] which can process massive amounts of structured and unstructured data and develop varied models and algorithms to find patterns of anomalies and predict fraudulent customer behaviour.

#### *Big Data and Sentiment Analysis*

Sentiment Analysis is the most extensively used application of big data. Presently, gazillions of conversations occur on the social media, which when connected to one's advantage can aid any business in determining new patterns, protecting image of their brand and segmenting consumer base to improve product marketing and customer experience. Several giants of the industry have presently developed tools for efficient sentiment analysis.

#### *Food Industry*

The impact of Big Data on the food industry [13] is increasing exponentially. It is applicable in tracking the quality of products, developing marketing strategies for better customer experience, presenting approvals to the customer, the presence of Big Data in the food industry is slowly becoming ubiquitous.

#### *Telecom Industry*

Concepts of Big Data and Machine Learning are gradually implemented for improving customer service and satisfaction. Call detail records, customer service logs, emails to social media, web logs, and geospatial and weather data are few examples of data being available to telecom operators. Handling huge amounts of data can be a scary task. Developing deep understandings with the support of Machine Language running on Apache Hadoop can help workers to economically take advantage of ever-increasing datasets so as to enhance their service quality and customer experience as well as to increase the customer base. There are immense benefits of using such technologies. Predictive maintenance ensures prediction, prevention and recovery of operational disruptions. Real-time processed data can be used to dynamically allocate the bandwidth to reduce congestion.

## FUTURE SCOPE AND DEVELOPMENT

Today, Big Data is influencing IT industry. The huge amount of data generated from mobile devices, sensor-enabled machines, cloud computing, social media and satellites help different organizations to improve their decision making practices and take their businesses to another higher level. Data is the biggest thing that hit the industry since personal computer was invented by Steve Jobs. As mentioned earlier, data is generated every day in such a rapid manner that, traditional databases will gradually give up in storing, retrieving, and searching some particular records from stored data. Big data technologies have addressed the problems related to big data revolution through the use of commodity hardware and distribution. Companies like Google, Yahoo!, Microsoft, Facebook, Amazon, and General Electric etc. are investing a lot in Big Data research and projects. Google has launched the Google Cloud Platform, which offers developers to develop a variety of products from websites to complex applications.

## REFERENCES

- [1] <http://hive.apache.org>
- [2] <http://blogs.worldbank.org/voices/meet-winners-and-finalists-firstwbg-big-data-innovation-challenge>
- [3] <http://www.internetlivestats.com/twitter-statistics/>
- [4] <http://www.internetlivestats.com/google-search-statistics/>
- [5] Grand Challenge: Applying Regulatory Science and Big Data to Improve Medical Device Innovation, Arthur G. Erdman\*, Daniel F. Keefe, Senior Member, IEEE, and Randall Schiestl, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 60, NO. 3, MARCH 2013
- [6] <http://www.forbes.com/sites/bernardmarr/2015/04/21/how-bigdata-is-changing-healthcare/>
- [7] <http://www.engadget.com/2015/04/10/jawbone-up3-shipping-april20th/>
- [8] <http://www.samsung.com/uk/consumer/mobiledevices/wearables/gear/SM-R3500ZKABTU>
- [9] <http://www.wsj.com/articles/how-computers-trawl-a-sea-of-datafor-stock-picks-1427941801>
- [10] Nitish Sinha, —Using Big Data in Finance: Example of sentiment extraction from news articles!; FEDS notes, March 2014
- [11] Baker, Malcolm and Jeffrey Wurgler, 2007. "Investor Sentiment in the Stock Market", Journal of Economic Perspectives, vol. 21(2), pages 129-152.
- [12] <http://www.pactera.com/resources/blog/how-big-data-isrevolutionizing-fraud-detection-in-financial-services/>
- [13] <https://datafloq.com/read/big-datas-impact-food-industry/96>