

UNDERSTANDING APPLICATIONS OF BIG DATA ANALYTICS THROUGH CASE STUDIES

Neeru Mago

Dept. of Computer Science and Applications, Panjab University

Swami Sarvanand Giri Regional Centre, Hoshiarpur, India

ABSTRACT

In this era of datafication, everything revolves around data which makes us possible to view, monitor, record and analyze everything as quantifiable data. Every day, we create 2.5 quintillion bytes of data. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals etc. Such colossal amount of data that is being produced continuously is what can be coined as Big Data. The term "Big Data" refers to the use of various advanced data analytics methods that extract value from data. Big data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. Hence it is a broad term for data sets which are extremely large or complex for which traditional data processing applications are inadequate. It includes various challenges like data analysis, capture, search, sharing, storage, transfer, visualization, and information privacy. In this paper, The 3V's of Big Data, Big Data Analytics, Application of Big Data analytics, Big Data Technology and various Big Data Case Studies are presented. In the end, future scope and developments are also discussed.

Keywords: *Big Data, Big Data analytics, Variety, Velocity, Volume,.*

I. INTRODUCTION

'Big Data' is a term used to describe collection of data that is huge in size and yet growing exponentially with time. In short, such a data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently. Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves as shown in Fig 1.



Fig 1: Big Data for decision and strategy making

The basic requirements for working with big data are the same as the requirements for working with datasets of any size. However, the massive scale, the speed of ingesting and processing, and the characteristics of the data that must be dealt with at each stage of the process present significant new challenges when designing solutions. The goal of most big data systems is to surface insights and connections from large volumes of heterogeneous data that would not be possible using conventional methods. In 2001, Gartner's Doug Laney first presented what became known as the "three Vs of big data" to describe some of the characteristics that make big data different from other data processing.

II CHARACTERISTICS OF BIG DATA

1. The 3V's of Big Data:

- **Volume** – The quantity of data that is generated is very important in this context. It is the size of the data which determines the value and potential of the data under consideration and whether it can actually be considered Big Data or not. The name 'Big Data' itself contains a term which is related to size and hence the characteristic.
- **Variety** – Big Data isn't just numbers, dates, and strings. Big Data is structured & unstructured geospatial data, 3D data, audio and video, sequences, time series, multi-dim arrays and unstructured text, including log files and social media, etc. To extract knowledge all these types of data need to be linked together.
- **Velocity** – The term 'velocity' in the context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development. Data is being generated fast and needs to be processed fast. Late decisions lead to missing opportunities.

2. Other important characteristics of big data

- **Variability** - This is a factor which can be a problem for those who analyse the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.
- **Veracity** - The quality of the data being captured can vary greatly. Accuracy of analysis depends on the veracity of the source data.
- **Complexity** - Data management can become a very complex process, especially when large volumes of data come from multiple sources. These data need to be linked, connected and correlated in order to be able to grasp the information that is supposed to be conveyed by these data. This situation, is therefore, termed as the 'complexity' of Big Data.

III BIG DATA ANALYTICS

Big Data analytics is the process of collecting, organizing and analyzing large sets of data (Big Data) to discover patterns and other useful information. Big Data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business

IV APPLICATION OF BIG DATA ANALYTICS

- **Government:** The use and adoption of big data within governmental processes allows efficiencies in terms of cost, productivity, and innovation.
- **International development:** Research on the effective usage of information and communication technologies for development (also known as ICT4D) suggests that big data technology can make important contributions but also present unique challenges to International development. Advancements in big data analysis offer cost-effective opportunities to improve decision-making in critical development areas such as health care, employment, economic productivity, crime, security, and natural disaster and resource management.
- **Manufacturing:** Based on TCS 2013 Global Trend Study, improvements in supply planning and product quality provide the greatest benefit of big data for manufacturing. Big data provides an infrastructure for transparency in manufacturing industry, which is the ability to unravel uncertainties such as inconsistent component performance and availability. Predictive manufacturing as an applicable approach toward near-zero downtime and transparency requires vast amount of data and advanced prediction tools for a systematic process of data into useful information.
- **Healthcare :** Big data analytics has helped healthcare improve by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions.
- **Education:** A McKinsey Global Institute study found a shortage of 1.5 million highly trained data professionals and managers and a number of universities including University of Tennessee and UC Berkeley, have created masters programs to meet this demand. Private bootcamps have also developed programs to meet that demand, including free programs like The Data Incubator or paid programs like General Assembly.
- **Internet of Things (IoT):**Big data and the IoT work in conjunction. Data extracted from IoT devices provides a mapping of device interconnectivity. Such mappings have been used by the media industry, companies and governments to more accurately target their audience and increase media efficiency. IoT is also increasingly adopted as a means of gathering sensory data, and this sensory data has been used in medical and manufacturing contexts.
- **Information Technology :** Especially since 2015, big data has come to prominence within Business Operations as a tool to help employees work more efficiently and streamline the collection and distribution of Information Technology (IT). The use of big data to resolve IT and data collection issues within an enterprise is called IT Operations Analytics (ITOA). By applying big data principles into the concepts of machine intelligence and deep computing, IT departments can predict potential issues and move to provide solutions before the problems even happen.

V BIG DATA CASE STUDIES

The case studies or Use cases of Big Data are as follows as shown in Fig 4.



Fig 4: Use cases of Big Data

1. Big Data for Customer Sentiment Analysis



How Delta Airlines uses Big Data for Customer Sentiment Analysis-

- Large airlines like Delta, monitors tweets to find out how their customers feel about delays, upgrades , in-flight entertainment , and more.
- For example, when a customer tweets negatively about his lost baggage with the airline prior to boarding his connecting flight. The airline identifies such negative tweets and forwards to their support team.
- The support team sends a representative to the passengers destination presenting him a free first class upgrade ticket on his return along with the information about the tracked baggage promising to deliver it as soon as he or she steps out of the plane.
- The customer tweets like a happy camper rest of his trip helping the airlines build positive brand recognition.

2. Big Data for Behavioural analytics

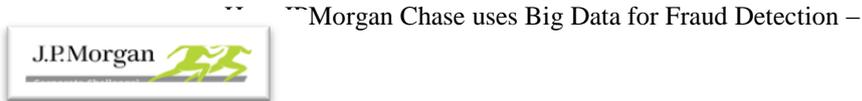


How McDonald's uses big data for behavioural analytics -

- With more than 34K local restaurants serving 69 million customers across 118 countries , 62 million daily customer traffic, selling 75 burgers every second, \$27 billion annual revenue- McDonald's is using big data analytics to gain lot more insight to improve operations at its various stores and enhance customer experience.

- McDonald's analytics system analyse data about various factors such as wait times, information on the menu, the size of the orders, ordering patterns of the customers to optimize the operations of its restaurants at specific locations.

3. Big Data for Fraud Detection



- JPMorgan Chase analyses emails, phone calls, transaction data to detect the possibilities of frauds which would otherwise be difficult to detect.
- JPMorgan uses analytics software developed by Palantir to keep a track of employee communications to identify any indications of internal fraud.

4. Big Data for Customer Segmentation



How Amazon uses big data for Customer Segmentation-

- Just have a look at Amazon and see what products it recommends you to buy. The products recommended by Amazon are probably different for you and your friend.
- How do they do it? Every time a user logs into his or her Amazon account and makes purchases or browses various products on the site, Amazon collects this data and the next time the customer's returns, they offer them products based on their previous purchase and browsing history.
- This also helps Amazon identify various trends amongst people who make similar purchases. For instance, if 75% of the people who buy an Apple iPhone 6s also buy a power bank, then Amazon offers power bank as a recommendation whenever somebody purchases an iPhone 6s.
- By segmenting the customers based on their interests and purchase patterns, Amazon provides people with more choices even if they are not looking to buy other products, thereby tempting them to make additional purchases.

5. Big Data for Predictive Support



How Ayasdi uses Big Data for Predictive Support-

Using 20 year old breast cancer data by running it through its IRIS product that combines machine learning and topology.

- Ayasdi is leveraging big data analytics to discover new relationships and new questions that can be answered with the help of the 20 year old dataset.
- Ayasdi developed topologies for leukaemia and breast cancer patients' data by analysing the data to find similarities which can help in predicting and finding novel cancer treatments and therapies.

VI FUTURE SCOPE AND DEVELOPMENT

- As far as the future of big data is concerned it is for certain that data volumes will continue to grow and the prime reason for that would be the drastic increment in the number of hand held devices and internet connected devices, which is expected to grow in an exponential order.
- Data are becoming the new raw material of business . Thus world is becoming data driven . Each and every decision are now taken on data from stock markets to machines that stalk you.

VII CONCLUSION

To handle big data and to work with it and obtaining benefits from it, a branch of science has come up and is evolving , called Data Science. Data Science is the branch of science that deals with discovering knowledge from huge sets of data, mostly unstructured and semi structured, by virtue of data inference and exploration. It's a revolution that's changing the world and finds application across various industries like finance, retail, healthcare, manufacturing, sports and communication. Search engine and digital marketing companies like Google, networking companies like Facebook and finance and e-commerce companies like Amazon are requiring and will require a lots of data scientists. As far as security is concerned the existing technologies are promising to evolve as newer vulnerabilities to big data arise and the need for securing them increases.

REFERENCES

- [1] Mukherjee et. al. " Big Data – Concepts, Applications, Challenges and Future Scope", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016.
- [2]<https://www.dezyre.com/article/5-big-data-use-cases-how-companies-use-big-data/155>
- [3]<http://whatis.techtarget.com/definition/fast-data>
- [4] Arif et.al. "Challenges and Security Issues in Big Data Analysis", IJRSET. Volume: 5. Issue: 1. January 2016.
- [5] Arthur et.al. "Grand Challenge: Applying Regulatory Science and Big Data to Improve Medical Device Innovation", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 60, NO. 3, MARCH 2013
- [6] Big Data Survey Research Briefl, SAS, The power to know, 2013
- [7] David Loshin, "Addressing Five Emerging Challenges Of Big Data", Progress Software, 2014
- [8] "Data Connectivity Outlook paper", Progress Software, January 2014.
- [9] Puneet Singh Duggal, Sanchita Paul ,— Big Data Analysis: Challenges and Solutionsl, International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV.
- [10] S. Ghemawat, H. Gobiuff, and S. Leung, "The Google File System" in ACM Symposium on Operating Systems Principles, Lake George, NY, Oct 2003, pp. 29 – 43.
- [11] Jefry Dean and Sanjay Ghemwat, "MapReduce: A Flexible Data Processing Tool", Communications of the ACM, Volume 53, Issue.1,January 2010, pp 72-77.