Big Data: Evolution cum Revolution

[1]Bhavesh Sharma, [2]Prabhdeep Kaur, [3]Saurabh Kumar

[1]Engineer, Digital, Mindtree Limited MTW, Karnataka, India
[2]Lecturer, BETHEL International PU College, Karnataka, India
[3]Associate Director, Delivery Anchor-RCM, Mindtree Limited MTW, Karnataka, India

---***---

Abstract - *Big Data has almost occupied every field, but without knowing how to extract and implement the data into an application, it is totally useless no matter even if we have petabytes of data. The magnitude of data is rising abruptly every year and if this data is set into its optimum use then the same data can enhance the decision-making process. This paper communicates about the evolution of big data and its applications in our daily life, it also emphases on the areas in which big data is helpful and how the data can be utilized effectively and efficiently. In this document you will get to know how we can convert massive amount of data into information which can be highly beneficial for an individual or an organization to achieve the desired throughput.*

Key Words: Big Data, Analytics, Insight, Decision Making, Data Warehouse, Machine learning, Artificial Intelligence, Internet of Things, Deep Learning, Hadoop, Map Reduce, Spark.

1. INTRODUCTION

Big data plays a vital role in our daily life and occupies almost every field in today's digital world. It is imperative to learn how the four Vs of data affect the business and daily applications and how we can use the big data analytics in various day to day applications like online shopping, business strategies, decision making etc. We can increase the efficacy of data usage if we come to know the processes involved in the extraction and management of big data. [1]

1.1 Data

Any text, sounds, videos, images and any type of files are considered as data. Data is gathered through social media, google search, blogs, e-commerce etc. [2,3]

1.2 Information

It is a cluster of data. From this cluster, we can get to know about an area. It means the data which makes sense. This step forms the base of big data abstraction. [4]

1.3 Analytics

Analytics is the process of transfer of data/information into useful material to take fruitful decisions in an oriented way. It is the major step which is used to take decisions. [5]

1.4 Insight

It is the inference drawn after performing analytics. The final step for decision making is insight. For example: You are working in a company and you are going to hire skilled employees for your company. In that case, you can analyze the employee from the information provided by the employee. So, only your insight into that person will be the deciding factor for his recruitment.

2. Big Data Applications

Big means giant and data is anything like text, sounds, videos, images or any type of file. So Big data is huge data and there is a massive problem to store this data. We can store the data in file formats, using VSAM and ISAM.

Data is increasing brusquely. Per day millions of Tbs of data is generated and it is very difficult to manage all the data. Data is the strength for every organization. In Big Data, we convert the data into Information then information is converted into Knowledge and then Knowledge is converted into Insight and then Insight is converted into wisdom and through wisdom, we can take impactful decisions for business purposes as well as for general works in our daily life. [6]

The big data includes the information generated by various gadgets and applications. Some of the areas which form the Octopus in Big data are:

1. **Health data:** collects lots of information related to health like step count, heartbeat and diseases.
2. **Search engine data:** Search engine extracts lots of user data through key words.
3. **Ola/Uber Data:** They collect your data through your login credentials and location.
4. **Stock Exchange Data:** It is an information of a customer when he purchases or sells the share and

how much profit he gets. Many companies/organizations are using big data analytics to buy shares of other companies.

5. **Social Media Data:** It's rising tremendously. Social media sites like (Facebook (FB), Twitter, Instagram, Snapchat and LinkedIn) carry lots of data like LinkedIn carries Jobs related data, Facebook and Instagram carries photos and videos, etc.

6. **Aadhar card:** It holds personal information for Indian citizens. [7]

3. Big data analytics in 2020

Big data analytics is the process in which we analyze the different forms of data types as the gigantic data is increasing day by day and due to which the volume of data is also rising exponentially. Examination of big data supports associations with furnishing them the information and they are using the same information to create new probabilities. It also increases the level of decision making, which gives a push to organizations'/companies' growth and further it is used to increase the profit and new customers. Business intelligence helps associations with better comprehension of the data which is used to extract the information and similarly used to differentiate the information which will be an asset to decide the present and future business scenarios. Professionals who are using large data/information needs the insight that will come by segregating the information. [8]

Analytics in Big data depends on three approaches. They are:

1. **Prescriptive:** In this approach, it will check what needs to be done. It involves:
   - Optimization
   - Decision trees
   - Mathematical programming
2. **Predictive:** In this approach, it will check what the future consequences and tendencies. It involves:
   - Forecasting
   - Data mining
   - Simulation
3. **Descriptive:** In this approach, it will check what is currently running. For instance:
   - Business intelligence
   - Dashboard. [9]

4. Importance of big data analytics in 2020

Big data analytics helps in organizations to control their data and use it to take intelligent business decisions and through which they can get higher profits. Big data analytics helps us to save lots of time and efforts. Many e-commerce websites are using big data analytics to understand the customer

requirement. They analyze the customer's needs and interests from previous purchases and they start showing the products which the customer needs and by this they will increase the profit. Many companies and organizations are shifting towards big data (Hadoop) as it is the best option to store the data efficiently and moreover it is cost effective. Companies/Organizations are using big data analytics as it is very fast, so it is very easy for the companies to take impulsive decisions proficiently.

Fig.1 Big Data Analytics: What is it and Why it matters? [4]

Through big data analytics, companies get to know the customer's requirements. So, they will make the products which customer needs. E-commerce websites use the customer's purchase history, most searched items, browsing history etc. to hype the sale of their products. Example: Amazon and Flipkart are using previous purchase and browsing history to do analytics on what majority of the customers had purchased and then based on collected data it automatically recommend products to the customer. Amazon and Flipkart are also tracking the quantity of products as you can see sometime messages like "only 1 product left". [10]

Nowadays most of the companies are trying to shift towards the streaming of Music, TV series and movies like Amazon prime video, Flipkart video, Voot app and Netflix. These companies/Organizations are using Big data to create new TV series and Short Movies. They collect data from user's

previous watch history and history of the browser (Which type of movies and tv series they have searched) from this they plan to create some impactful new tv series. Netflix and Amazon prime video collect the data and find which actors,

actresses, directors and skits are famous, then they use this data and create new TV shows and movies to earn profit by making TV shows and movies by following that trend. Many customers see the rating of the shows to take decision if it is worthy to watch this movie or not. [11]

The best point about big data is we can fight with crime. For instance, we can collect all the crime related data and find out which state or city has the highest crime rate and at what time. So, it will help the police department to know the place and time of occurrence of any new crime. Here they can use descriptive, prescriptive and predictive analysis to find out the place and the time of new crime scene from this data and they can stop the crime before it happens.

Big data also helps doctors to understand the patient's diseases through symptoms. We are using big data to measure the heartbeat of the body and to calculate the footsteps, the calories and many more. So, it is obvious that big data is making an impact in almost every field. If careful analysis is being done, big data can prove to be of great help to every person. [12]

5. Challenges faced by Big Data till 2020

Every evolution comes with certain challenges and so is the big data. As the fields involving usage of big data are multiplying, the challenges are also coming into picture. Some of the challenges faced by Big data are:

1. Collecting the data
2. Fetching the data
3. Analysis of data
4. Acquisition of professional skills to use big data
5. Searching the data/information
6. Data integration
7. Volume of data
8. Sharing of data
9. Presentation
10. Velocity
11. Veracity
12. Variety
13. Security

It is imperative to carefully visualize the challenges and make them incompetent to restrict the usage of big data in its optimum way.

6. Technologies using Big data till 2020

Deep learning: Deep learning is a deep structured learning and it is a branch of AI, detailed diagram challenges the complex abstractions in information by collection of algorithms with several processing films, unruffled by several linear and non-linear conversions. DL belongs to Artificial intelligence technique based upon diagrammatical representation of data. A graphical representation (e.g. a diagram) may be characterized in different ways such as intensity values per pixel. There are various depictions which give a clearer view than others because of their easy representation (e.g. facial look identification and face identification). Earlier people used to struggle with the handcrafts as the representations were also vague but with the help of deep learning algorithms, this issue has been fixed. [13]

Machine learning: Machine learning is a part of Artificial intelligence (AI) and it gives power to the computer to learn without any external program. Machine learning focuses on the quality of pc program which learns from new data which can modify the program according to the same. The method of ML is analogous to the data mining technique. ML and data mining systems retrieve the data to find the patterns. ML algorithms can be considered as either supervised or unsupervised. [14]

- **Supervised learning**: In Supervised learning, the machine collects the historical data and perform functions on it to get desired output.

- **Unsupervised learning:** In Unsupervised learning, the machine learns from the data without any desired output. Not only it learns but also discovers the mysterious information.

Artificial intelligence: Artificial intelligence (AI) is a part of human intelligence, which is used to create software/computer/chat bot and many more that makes them to behave like human. For example: Alexa, google mini, etc.

Artificial intelligence includes the following areas of specialization:

- **Expert system**: It creates a computer program which is used to take decision in actual life situations (for e.g. expert system helps doctors to detect diseases based on symptoms).

- **Neural networks**: It is the network which act as a human brain and stimulate intelligence by reproducing neurons and it learns from the dataset where instructions are already given.

- **Robotics**: It can be defined as the program which interacts with other programs to take the action.

For e.g. Weather control in car is considered a part of Robotics. So, in simple language, Robotics is a kind of machine program which is not controlled by a human being.

- **Natural language**: The program which recognizes natural human languages.

- **Games**: The program which is used to play against one on one with human mind. For instance, the chess program is proficient to defeat a human mind. In 1997, an IBM super-computer called deep blue defeated world chess champion Gary Kasparov in the second match of chess.

IoT (Internet of things): IoT is nothing but it's method in which systems are connected with each other through internet and Internet is used to pass the data/instruction to the other system without human interference. It uses the sensors and software to exchange the data over internet.

Cloud Computing: It is the virtual space which is used to store data online. It is used for creating, configuring and retrieving the data online. It offers online data storage, application and infrastructure. We do not need to install any software to install on our local PC. [15]

7. Importance of business intelligence in 2020

It plays an important role in decision making. As we know that insight is not enough to make business related decisions easily. For e.g. everyone has its own perspective towards problem, but every perspective cannot take virtuous decisions in an organization. So, to solve this problem Business intelligence term comes into the picture.

Business intelligence is defined as the sum of data warehouse and decision support system. Data warehouse is nothing but it's the place where data is collected and then organized. The organized data is stored into a database. Then the decision support system retrieves the collected/maintained data from data warehouse followed by selecting the required data from data warehouse and manipulating the data according to the requirement. After that the manipulated data is analyzed and presented in a file to take the final decisions.

The following steps are used in Data warehouse:

1. Collect
2. Transform
3. Organize

4. Store
5. Maintain

The steps used in DSS are:

1. Select
2. Manipulate
3. Analyze
4. Present

Business intelligence takes historical data or understand the past data and do analytics on that to optimize the future.

Tools and applications in Business intelligence

1. Business object
2. Cognos insight
3. Sap-BW

Value of information: Information plays an important role to take decision. Nearly 80% of employees' time is spent to find the important information. Many business challenges will be solved through the information.

- It has capability to make fast and good decisions to optimize business process.
- Accurate picture of customers' needs improves the customer service.
- It will take minimal time for searching the optimum solution and then ultimately increase the employee productivity.
- It is transparent, has capability to reduce risks. [16]

8. Innovations in Big data Technology till 2020

In 2005 the data started to rise tremendously and the name "Big data" was introduced and it was defined as a huge data impossible to store, manage and process using past technologies. Before 2005 Google launched the map reduce technique to structure the data. In 2005, yahoo also introduced Hadoop which was created on top of the MapReduce. Yahoo created Hadoop to store all the websites' names (www) in the form of index. After that searching became easy and fast. Now many organizations are using Hadoop to store data in HDFS (Hadoop Distributed File System). Basically, Hadoop is an environment in which HDFS and MapReduce both are used to do analytics. More than 95% of the data which is available now, has been created in the last 3 years. According to research, the data created in one year in the year 2011 is equivalent to the data created in 1 day in 2020. In 2019, 45 percent of companies had shifted to Big data. In 2019 and 2020 lots of Bigdata startups have also been started. Now companies are trying to learn big data and preparing themselves to shift to big data. Spark is

nothing but a framework with additional features on top of Hadoop. Nowadays no one is using Hadoop, everyone is shifting to Spark because spark is 99 times quicker than map reduce jobs and spark codes can be written using programming languages like Java, Scala, python and R. [18]

The tools which are used in Big data are:

1. Spark 2.4(enhanced version of Hadoop)
2. Pig
3. Hive
4. HBase
5. Zookeeper
6. Kafka
7. Yarn
8. Cluster manager
9. Oozie

Conclusion

Even a small amount of data can prove to be an asset. So, one can understand how domineering is the Big data for an organization. Big data has brought up a revolution in almost every field whether it is related to health, marketing, entertainment or any other field involving the usage of data or information in one or the other way. Understanding the concept of big data and its efficient usage to increase the productivity has become a matter of concern for every small scale and large-scale organization. Through this paper, we have tried to cover the various aspects related to management and extraction of big data to achieve the decision-making goals in an organization. Big data if used in a proficient manner can change not only the present as well as the future scenario of a specific organization. A careful analysis, intellectual extraction and competent decision making by using big data can bring laurels to an organization because data is the only bow to shoot the arrow of success of a skilled organization. Big data being the base of today's digital world has been a radical evolution and with its effects slowly turning into a revolution in almost every sphere of life.

REFERENCES

[1] J.G Lee, M. Kang, "Geospatial Big Data: Challenges and Opportunities", vol. 2, Issue 2, Jun. 2015, pp.74-81.

[2] M.Kowalczyk, P.Buxmann, "Big Data and Information Processing in Organizational Decision Processes", Aug.2014, pp.267–278.

[3]. A. Chakrabarti, S. Parthasarathy, C. Stewart, "Green- and heterogeneity-aware partitioning for data analytics", Proc. of IEEE Conference on Computer Communications Workshops, 2016, pp. 366–371.

[4]. "Big Data Analytics: What is it and Why it matters?", [Image Source ONLINE] Available at https://www.sas.com/en_us/insights/analytics/big-data-analytics.html, [Accessed on 3rd March,2020]

[5]. S. Salloum, R. Dautov, X. Chen, P. X. Peng, and J. Z. Huang, "Big data analytics on apache spark", International Journal of Data Science and Analytics, vol. 1, no. 3,2016, pp. 145–164.

[6]. S. Salloum, J. Z. Huang, and Y. He, "Exploring and cleaning big data with random sample data blocks", Journal of Big Data, vol. 6, no. 1, 2019, pp. 45.

[7]. Top 10 Applications of Big Data Across Industries, [ONLINE] Available at: https://www.simplilearn.com/ [Accessed on 6th March, 2020].

[8]. K. Sravanthi, T.S. Reddy, "Applications of Big data in Various Fields", International Journal of Computer Science and Information Technologies, vol. 6, issue 5, 2015, pp.4629-4632.

[9]. A. Banerjee, "Big data and advanced analytics in Telecom: A Multi-Billion-Dollar Revenue Opportunity," Dec. 2013.

[10]. "The Four V's of Big Data", [ONLINE] Available at: http://www.ibmbigdatahub.com/infographic/four-vs-big-data. [Accessed on 3rd march, 2020].

[11]. G. Linden, B. Smith, J. York, "Amazon.com Recommendations Item to Item Collaborative Filtering", IEEE, vol.3, issue 1,2003, pp.76-79.

[12]. G.Hughes, "How big is 'big data'in healthcare?" [ONLINE] Available at: http://blogs.sas.com/content/hls/2011/10/21/how-big-is-big-data-in-healthcare/. [Accessed on 7th March,2020].

[13]. " Global Artificial Intelligence (AI) Market is valued at US\$ 900 million by year end 2013", [ONLINE] Available at: http://beforeitsnews.com/science-and-technology/2013/04/global-artificial-intelligence-ai-market-is-valued-at-us-900- million-by-year-end-2013-2569204.html. [Accessed on 7th March,2020].

[14]. A. Rajaraman, J. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2011.

[15]. M. Armbrust, A. Fox, R. Griffith, A. Joseph. R. Katz, A. Konwinski., G. Lee, M. Zaharia, "A view of cloud computing", Communications of the ACM, vol.53, issue 4, Apr.2010, pp.50-58.

[16]. R. Polock, "Forget Big Data, Small Data is the Real Revolution", [ONLINE] Available at: https://blog.okfn.org/2013/04/22/forget-big-data-small-data-is-the-real-revolution/ [Accessed on 8th March,2020].