AN INSIGHT INTO THE USAGE OF BIG DATA ANALYTICS

Chandralekha M
Research Scholar
Department of Computer Applications
University College of Engineering
Bharathidasan Institute of Technology (BIT) Campus
Anna University, Tiruchirappalli - 620 024

Dr. N. Shenbagavadivu
Assistant Professor
Department of Computer Applications
University College of Engineering
Bharathidasan Institute of Technology (BIT) Campus
Anna University, Tiruchirappalli - 620 024

Abstract: Big data analytics is pronounced in a profound manner among all. Elaborate discussions are explored explicitly made regarding data generation, collection, usage, archives, types and characteristics deployed in all along the big data spectrum. Various trends are analysed in different fields of applications in big data. As the entire world is marching towards the big data, special attention is paid by every data scientists to extract the benefits out of it. The objective of this paper is to focus the future utility of big data vividly.

*Keywords:*Big data analytics; Data scientists; Smart Grid; E-health; IoT; Utilities; Logistics

1. INTRODUCTION

Today, the amount of data is exploding at an unprecedentedrate as a result of developments in Web technologies, socialmedia, mobile and sensing devices. For example, Twitterprocesses over 70M tweets per day, thereby generating over8TB daily [1]. ABI Research estimates that by 2020, there will be more than 30 billion connected devices [2]. TheseBig Data possess tremendous potential in terms of businessvalue in a variety of fields such as health care, biology,transportation, online advertising, energy management, and financial services [3], [4]. However, traditional approachesare struggling when faced with these massive data.

Mammoth datais playing a vital role in all forthcoming events. Different types of data are being generated from various sources such as health, government, social networks, marketing, financial etc. on daily basis. As the quantum of data generation is large, it is essential to focus more on data collection. At the advent of emerging technologies like Cloud Computing, Internet of Things (IoT), electronic gadgets,one has to pay more attention in the generation,storage and archive of data. Predominantly, distributed data are widely used in the areas like smart grid systems, healthcare systems, retailing systems like that of Walmart, government systems etc.

Earlier, handling huge data has become a challenge to the human. Inorder to overcome the issues like storing, processing, managing and archiving, Big Data has been found to fill that void.

The remainder of this paper is organized as follows:Section II defines big data, and Section III presentsBigData applications. Finally, Section IVconcludes the paper.

2. BIG DATA DEFINITION

Big data consists of datasets with the format of structured, unstructured and semi-structured data. Handling of burgeoning data is hectic in nature as it needs latest state-of-the-art technological background as well as complex algorithms. Usages of Conventional tools become invalid in big data applications.

Data experts characterized big data on the basis of volume, velocity and variety (called as 3Vs) [5]. Fig 1 represents the 3 V's of Big Data.
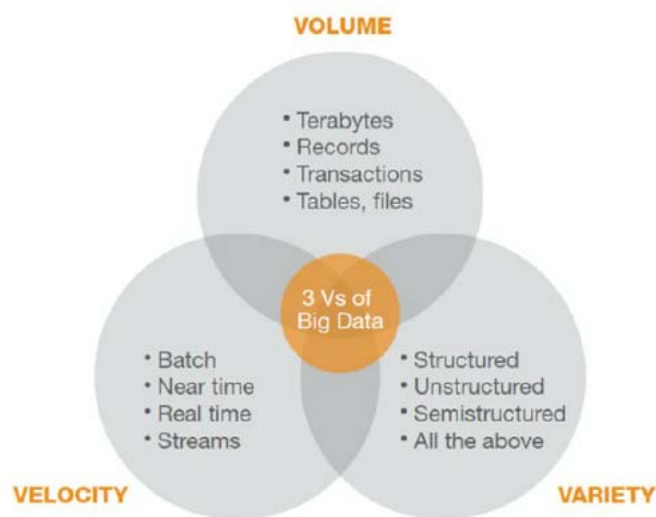

Fig 1: The 3 V's of Big Data

Volume:Data generation is taking place regularly from innumerable number of devices such as ICTs, smartphones, products' codes, social networks, sensors, logs, etc. As per McAfee et al. (2012) [6],the data generation per day was 2.5 Exabyte in 2012. This volume became doubled in every 40 months. International Data Corporation (IDC), a company which publishes research reports revealed 4.4 Zettabytes (ZB) data generation in 2013. The above quantum was doubled in every two years. The data generation became 8

ZB in the year 2015 as stated by Rajaraman (2016) [7]. Accordingly, IDC predicted that the escalation of data generation in the year 2020shall be upto 40 Zeta bytes which are around 400 times as stated by Kune et al., 2016 [8].

Velocity:The rate at which the data generation shall be accrued in an enormous manner on day today basis in almost all sectors. It is a hectic process to convert the same into useful information. The data generation in Wallmart (an international discount retail chain) is greater than 2.5 PB/hour among the transactions done by the customer. Similarly, YouTube is also generating big data in a speedy manner.

Variety:Different types of data like videos, documents, comments, logs etc. are being generated from various sources such as structured and unstructured data, public or private, local or distant, shared or confidential, complete or incomplete, etc.

As per Emani et al. (2015) and Gandomi and Haider (2015) [9], [10], the above 3 Vs are further expanded with the inclusion of **Vision**which is used for a specific purpose; **Verification** is done with respect to some specifications; **Validation** is to fulfil the purpose and **Value**is taken from any raw data from the field to process into micro level. In addition, complexity and immutability are taken into account to well manage the data. Fig 2 represents Big Data 5Vs.
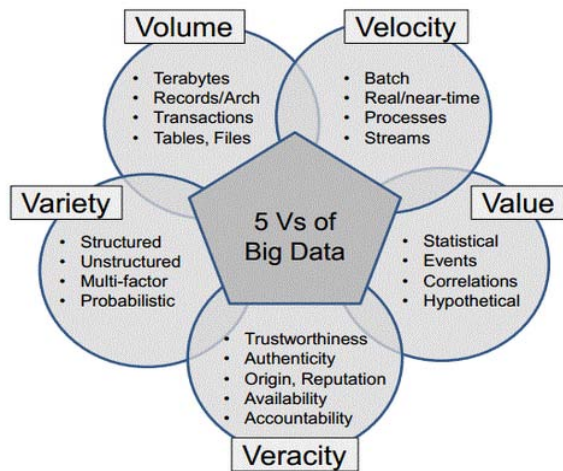

Fig: 2Big Data 5Vs

3. BIG DATA APPLICATIONS

Big data applications are defined as below:

1. **Power Management in Smart Grid:**It is important to manage power in smart grid. Power generation takes place in various sources like Thermal Power Station, Atomic Power Station, Hydro Power Station and Nuclear Power Station. The generated power is being transmitted to the various parts of the countries through various corridors by engaging electrical hardware like Transformers, Switch gears, Sensors, Monitors through smart grid. In order to handle smart grid operation effectively and proper optimisation, the data collected across the smart grid is very helpful to predict and match the supply demand and monitor and control the healthiness of the system in a cost effective manner.

2. **E-health:**Population in India needs to pay more attention particularly in the health. Our medical facilities are inadequate to reach the public hailing from urban areas. This issue can be eliminated by using technology like Internet. Health department covers right from clinical data, pharmaceutical data, laboratory data, instrument data and hospital management. In depth analysis of medical datasets has plenty of useful applications enabling to personalize the health service. For example, physician shall monitor patients' history through online to fix the prescriptions and diagnose accordingly. Data involved in healthcare systems shall not only optimise the hospital operations but also to minimise the health cost.

3. **Internet of Things (IoT):** The Internet of things (IoT) is the network of physical devices, vehicles, home appliances, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to connect and exchange data. Each thing is uniquely identifiable through its embedded computing system but is able to inter-operate within the existing Internet infrastructure. Experts estimate that the IoT will consist of about 30 billion objects by 2020. The IoT allows objects to be sensed or controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems, and resulting in improved efficiency, accuracy and economic benefit in addition to reduced human intervention. When IoT is augmented with sensors and actuators, the technology becomes an instance of the more general class of cyber-physical systems, which also encompasses technologies such as smart grids, virtual power plants, smart homes, intelligent transportation and smart cities.Big data plays significant roles in logistics for sending goods from one place to another. A lot of vehicles, man power are being engaged to deliver things to its destination. To track the above data histories that is generated in a big form through wireless adapters, sensors and GPS. Big data is applied to monitor the entire operations and optimise the delivery routes to distribute in time effectively.

4) **Public Utilities:** Nowadays water distribution to the public has become inevitable and important particularly in metros. To monitor the entire water flow through pipes, leakage through pipes, flow sensors, illegal connections and quantification in billing, the data generation is useful to incorporate the big data concepts. The maintenance throughout the water lines can be taken effectively with the help of employing the big data availability in the control system.

5) **Transportation and Logistics:**Big data is also useful in road transport systems. Lakhs of buses are being operated throughout the state and inter-state to commute the passengers from one place to another. Thousands of employees and buses are being operated on daily basis. Tickets are issued to the public to permit the passengers to reach their destination. The entire operation history are generated through big data is useful to monitor, maintain and optimise the system effectively. Online reservation is

also made possible with the help of internet. Similarly, Indian railways are one of the biggest railway departments in the world. Lakhs of trains are operated across the country. Lakhs of people are employed in Indian railways to maintain and operate the systems. A computerized reservation terminal is available in almost all stations to book the tickets in advanced manner. The data collection, generation in the railway is big data indeed.

4. CONCLUSION

As a result of discussing the various applications of big data in various segments, the complex collection of data generation, data archive and data storage are carried out effectively by using various datasets and advanced algorithms. Big data is quiet useful in futuristic technological breakthrough envisaged in and around the world. The previous data trendsuseful to the human to generate the predictive model for implementing the process to move forward in an optimised manner.

5. REFERENCES

[1] R. Krikorian. (2010). Twitter by the Numbers, Twitter.[Online].
Available:http://www.slideshare.net/raf_krikorian/twitter-by-the-numbers?ref=http://techcrunch.com/2010/09/17/twitter-seeing-6-billion-api-callsper-day-70k-per-second/

[2] ABI. (2013). Billion Devices Will Wirelessly Connect to theInternet of Everything in 2020, ABI Research. [Online]. Available:https://www.abiresearch.com/press/more-than-30-billion-devices-willwirelessly-conne/

[3] W. Raghupathi and V. Raghupathi, ``Big data analytics in healthcare:Promise and potential,'' Health Inf. Sci. Syst., vol. 2, no. 1, pp. 1-10, 2014.

[4] O.Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha, ``Efficient machine learning for big data: A review,'' Big Data Res., vol. 2,no. 3, pp. 87-93, Sep. 2015.

[5] Furht, B., Villanustre, F., 2016. Introduction to big data. In: Big Data Technol. App..Springer International Publishing, Cham, pp. 3–11.

[6] McAfee, A., Brynjolfsson, E., et al., 2012. Big Data: the management revolution.Harvard Bus. Rev. 90, 60–68.

[7] Rajaraman, V., 2016. Big data analytics. Resonance 21, 695–716

[8] Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R., Buyya, R., 2016. The anatomyof big data computing. Software: Pract. Experience 46, 79–105.

[9] Emani, C.K., Cullot, N., Nicolle, C., 2015. Understandable big data: a survey. Comput.Sci. Rev. 17, 70–81.

[10] Gandomi, A., Haider, M., 2015. Beyond the hype: big data concepts, methods, andanalytics. Int. J. Inf. Manage. 35, 137–144.