A Survey on Big Data Issues and Challenges

*Shivaraj Koti ** Prof. Shivananda V.Seeri

Department of Computer Science, KLE Technological University, Karnataka, India
*** Department of MCA, KLE Technological University, Karnataka, India*
Corresponding Author: Shivaraj Koti

Abstract: In the last few years there has been rapid growth of data being generated or produced in various ICT (information and communication technology) based organizational processes. Such huge data is being generated by all kinds of sources and applications such as social network services, cloud services and knowledge bases. The generated data often in the form of structured, unstructured and semi structured format, and is often processed using computing platforms such as heterogeneous clouds, hybrid architectures and Hadoop. Using these techniques we deal with the real time data such as sensing data, e-commerce data, business transactions and web logs, medical data and etc. Big Data Analytics is a multi-application predictive data analytic platform designed as a scalable and extensible prediction engine for Big Data. It uses all available information (internal/ public) to generate predictions while modeling the entire population. Big Data Analytics is not a single technology, but a data-driven approach we use to develop and deploy customized solutions. Therefore big data analytics is the current area of research and development. The basic objective of this paper is to explore on big data challenges and research issues. It helps new researchers to develop solutions to the challenges and research issues.

I. Introduction

Big data can be defined as huge amount of data which requires new trending technologies and architectures to make possible to extract useful knowledge from it by doing some analysis work. New sources of data include data from government organization, private organization or personal devices such as mobile phones and desktops. The concept of big data rose because we leave in society which makes use of data intensive technologies [1]. Due to large amount of data it becomes very difficult to process and analyze the data using existing technologies [2]. Since big data is trending technology which helps in analyzing huge amount of data, which intern brings benefits to the many private and government organization. It is must that challenges and issues related to big data needs to be understood. The difficulties can be related to data capture, storage, search, sharing, analytics and visualization etc. Normally big data defined by 4 V's [1]. The 4 V's refers to velocity, volume, variety and veracity. The figure 1. Clearly describes about the characteristics of big data.

Characteristics of big data
A. Data volume
   There is tremendous amount of data being generated on every day, which strain on the conventional data storage of major companies and organizations. According to fortune magazine [9], up to 2003, the human population generated 5 Exabyte's of digital data. That amount of data was just produced in two days in 2011 and in 2013 we are generating that amount of data in just 10 minutes. The total amount data stored in computer today getting doubled every 12 to 18 months, that's the reason we live in the age of big data.
B. Data velocity
   To get the best result out of a big data, all the contents of a big data must be processed dynamically to generate quick results [3]. In the other side, velocity is the essence. With modern algorithms of big data analytics, it is now becoming easy to feed data into business processes and extract useful information from it.
C. Data variety
   One of the important characteristics of big data is variety of contents and sources involved, which opens up the new opportunities to create value from these variety of data [3]. The data which comes in different types of formats, some data may in structured format which are convent database entries, some it's may in semi-structured like images in metadata and the rest of is completely unstructured like text, graphics ,audio files, satellite imagery and streaming video.

D. Data veracity

The accuracy of matters a lot in any business decision.1 in 3 business leaders don't trust the information used to make the decisions [1]. Poor data quality costs many economy of government or private organization huge loss in terms of millions of dollars.
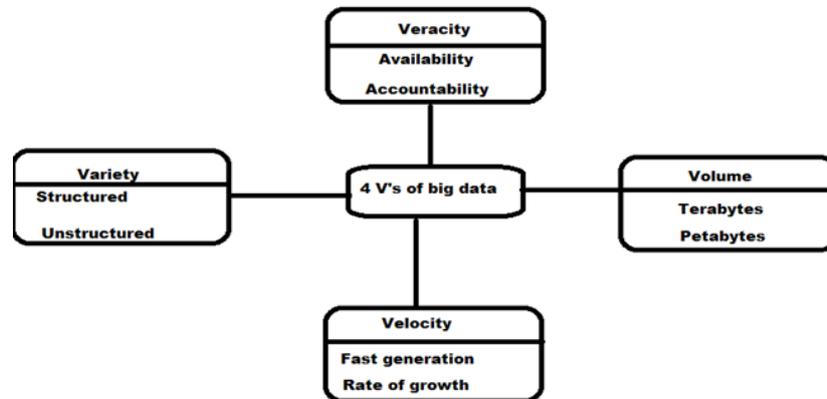

Fig. 1: Characteristics of big data

Issues in big data

Issues are the some of the conceptual points every company needs to know, to implement the technology effectively and efficiently. Big data issues need not be confused with the term problems, but they are crucial and we need to handle them.

A. Related to characteristics:

Data volume increases, the value of different data records will decreases in proportion to type, richness and quantity among other factors [3]. Say for many social networking sites producing data Peta bytes and it are increasing day by day, it is difficult to handle this amount of sheer data using traditional systems.

Data velocity, our traditional systems are not capable of doing analytics on data which is constantly in motion. Our e-commerce websites has rapidly increase the speed and richness of a data used for different business transactions [7]. In this case data velocity management more than the bandwidth issue.

Data variety, big data consists of different data which are raw, structured; semi-structured and unstructured which is difficult to handle by many existing analytical systems [3]. Because data contains incompatible data formats, non-aligned structures and improper data semantics posses significant challenge to analytical systems.

Data veracity should provide the confidence in the truthfulness of a data. But there is an uncertainty due to data inconsistency and incompleteness, ambiguities, latency and deception.

B. Transport and storage issue

New storage medium are invented due to quantity of data being exploded every time. Now a day's data is being created by everyone and everything i.e. from mobile devices to computers, not just by devices, by scientists, journalists, writers etc [3].

As we know huge amount of data being generated now day's inters of Peta bytes. Access to that data would overwhelm current communication networks. In this case it would take longer time to transmit the data, than to process it. To deal with such issue, they should be processed in the place and transmit the resulting or knowledge information. In other words, "bring the code to data", unlike the traditional method of "bring the data to code."

C. Management issue

It will be the most difficult problem to address with big data. The sources of data differ in much way by size, by format and method of collections. Many systems contribute data in the form of documents, videos, audios, drawings without the adequate information about the metadata from where, how, why, when and how it was collected [3]. Unlike in manually collecting data certain protocols are used to collect the data to ensure accuracy and validity. In simple way there is no perfect data management solution. This gap must be filled with new research techniques.

D. Processing issue

Rapid development of engineering applications like social network, sensor networks, semantic web and location based services; a variety of data has to be processed continuously to witness quick increase [4]. Effective and efficient management and processing of large-scale data poses an interesting and critical challenge. Recently, big data grasped the lot of attention from industry, from government and as well as government. These areas produces data in terms of Exabyte's of data, to process this much of data new parallel processing techniques and analytical algorithms are required.

Challenges in big data

Challenges are real implementation hurdles which require immediate attention. We need handle these challenges while implementing, otherwise it may lead to failure of the system and abnormal results.

A. Privacy and Security:

It is the most important challenge with big data which is sensitive and includes some conceptual and technical significance. When personal information of individual person when combined with external large data set, it leads inference about a person. Which may be secretive and person don't want their any information about them to be revealed [6]. Information regarding customer is collected and uses for business decision are made. This is done by creating insights into their lives and they are unaware of.

B. Data access and sharing information

The data in the company information is used to make business decision. It is important that data should be available in accurate manner, on time and complete. Expecting sharing of data between two companies or organisation is awkward. Sharing of data about their clients and operations threatens the culture of the company.

C. Analytical challenges

The type of analysis can do on the data which structures, un-structured or semi-structured which large amount of skills to extract useful knowledge for the best decision making [3]. This can be done by incorporating some best analysis methods.

D. Heterogeneous data

Unstructured data which is almost every kind of data being produced like social media sites, recorded meetings, fax report, pdf documents and many more. Working with unstructured data is costly. Converting unstructured data into structured data is not feasible. Structured data is always organized into highly mechanized way and manageable way.

E. Scalability

It is a characteristic of big data system that describes its capacity to cope and perform under an increased and expanding data. Scalability issue of big data has lead towards cloud computing, which now aggregates multiple disparate workloads with varying performance goals into very large clusters [3]. In recent years, there has been huge shift in the technologies being used. Hard disk drives are being replaced by phase change and solid state technology, which are not having the same performance sequential and random data transfer.

II. Conclusion

In a recent year we have seen sheer amount of data are being generated. To analyse these huge amount data is challenging for any common man. In this paper we have discussed various research issue and challenges of a big data. Here some of the important issues are covered that are needed to be analyzed by the organization while estimating the significance of implementing the Big Data technology and some direct challenges to the infrastructure of the technology. We believe that in future researchers pay more attentions to these issues and challenges for solving any big data problem effectively and efficiently.

References

[1].   D. P. Acharjya, Kauser Ahmed P, Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, School of Computing Science and Engineering, IT University Vellore, India, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016

[2].   World's data will grow by 50X in next decade, IDC study predicts. (2011, June 28) Retrieved from Computer World: http://www.computerworld.com/s/article/9217988/ World_s_data_will_grow_by_50x_in_next_decade_IDC_study_predicts.

[3].   Lenka Venkata Satyanarayan, Survey on Challenges and Advantages in Big Data, Dept. Of CSE, Aditya Institute of Technology and Management, Tekkali, AP, India, ICJCT Volume 6, Issue 2.

[4]. Changqing Ji,Yu Li,Wenming Qiu, Big Data Processing: Big Challenges and Opportunities, School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, Journal of Interconnection Networks Vol. 13, Nos. 3 & 4 (2012) 1250009 (19 pages).

[5]. Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Big Data: Survey, Technologies, Opportunities, and Challenges, Mobile Cloud Computing Research Lab, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia,

[6]. Ms. Vibhavari Chavan, Prof. Rajesh. N. Phursule, Survey Paper on Big Data, Department of Computer Engineering JSPM's Imperial College of Engineering and Research, Pune, Vibhavari Chavan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7932-7939.

[7]. Sabia and Love Arora, Technologies to Handle Big Data: A Survey, Department of Computer Science & Engineering, Guru Nanak Dev University, Regional Campus, Jalandhar, India.

[8]. Dilpreet Singh and Chandan K. Reddy, "A Survey on Platforms for Big Data Analytics", Department of Computer Science, Wayne State University, Detroit, MI 48202, USA Journal of Big Data, 1:1, 8, 2014.

[9]. Exponential rise in data volume, by fortune magazine study predicts, (2014, January) Retrieved from Thales group: https://www.thalesgroup.com/en/worldwide/four-vs-big-data.