A Study of Various Challenges of Big Data

Sanjeev Kumar*

Assistant Professor, Department of Computer Science, DAV College, Abohar, Punjab, India

Abstratct – Massive, quick and various knowledge moving quickly all over making what is referred to as "Big Data" era. This knowledge becomes vital source for valuable insights and ultimately serving to create a lot of informed call. But this knowledge with terribly special attributes can't be managed and processed by these ancient computer code systems, That became a true drawback. This study can discuss all different challenges of huge knowledge classified into 3 main groups: Data, method and management challenges. Knowledge challenges are the cluster of the challenges relates to the characteristics of the information itself. Method cluster includes all the challenges encountered whereas processing the massive Data; started with capture step and finished with presenting the output to purchasers

Keywords: Big Data, Hadoop, Scalability, Velocity and Volume.

- X -

1.   INTRODUCTION

1.1   Big Data

Big data could be a common phrase accustomed describes a huge volume of structured and unstructured knowledge that's therefore massive that it's tough to method with ancient information and package techniques. The characteristics that generally distinguish huge knowledge are the "3 V's" are volume, velocity and variety. In earlier a comparatively little volume of analog knowledge was created and created obtainable through a restricted range of channels, these days a huge quantity of information is frequently being generated and flowing from varied sources, through completely different channels, each minute in today's Digital Age. In fact, today, we tend to float in knowledge Ocean. During a broad range of application areas, knowledge is being collected at unexampled scale.

1.2   Challenges:

Big Data supply organizations with huge insight; but terabytes or pet bytes of knowledge flowing daily to a company have discovered that current infrastructures and architectures don't seem to be sufficient to satisfy the challenge. IT scientists are accountable to supply the technology capable of managing all technical needs of tremendous streams of knowledge. IT specialists are becoming a lot of calls as information grows; the requests are for a lot of Ad-Hoc analysis and summarized reports. Call manufacturers can't sit up for hours or days to seek out replies to queries if potential. These are samples of the challenges of massive information which may be classified into 3 main classes supported the

information life cycle: data, method and management challenges. Information Challenges are those pertain to the characteristics of the information itself, for instance information volume, variety, velocity, veracity, volatility, quality, discovery and intolerance. The second cluster is that the method challenges that are associated with series of however techniques: a way to capture information, a way to integrate information, a way to rework information, a way to choose the proper model for analysis and the way to supply the results. The third class is that the management challenges that cowl all privacy, security, governance and moral aspects.

2.   REVIEW OF LITERATURE

Shashi Shekhar et al [1] "Spatial big-data challenges crosswise over quality and distributed computing" more and that's only the tip of the iceberg, area aware datasets are of a size, variety, and update rate that exceeds the aptitude of spatial processing technologies. This paper addresses the rising challenges expose by such datasets, that we tend to decision spatial massive knowledge (SBD). SBD examples embody trajectories of cell-phones and GPS devices, vehicle engine measurements, temporally elaborate guides, etc. SBD can possibly remodel society by means of next-generation directing services like eco-steering. However, the envisaged SBD-based next-generation steering services create numerous crucial challenges for current directing techniques. SBD magnifies the effect of halfway information ANd ambiguity of ancient steering queries, for example, by a begin area and a complete area. also, SBD challenges

the idea that one algorithmic rule using a specific dataset is acceptable for all things.

Alexandru Adrian et al [2] "Big knowledge Challenges", the number of information that is traveling over the web these days, not solely that is mammoth, however is advanced also. Companies, establishments, healthcare system etc., every one of them use piles of information that are used for making reports in order to make sure coherence relating to the services that they need to supply. The method behind the results that these entities requests represents a challenge for computer code developers and firms that give IT infrastructure. The challenge is an approach to manipulate a powerful volume of information that must be immovably delivered through the web and reach its destination unblemished. This paper treats the challenges that massive knowledge creates.

Nrusimham Ammu Mohd rfanuddin et al [3] "Big knowledge Challenges" massive knowledge, An umbrella term for the explosion inside the sum and variety of high frequency advanced knowledge. The massive knowledge is additionally logs, mobile-keeping money exchanges; on-line user-generated content like diary posts and Tweets, on-line searches, satellite pictures, etc.- into uncalled for information needs exploitation process techniques to unveil trends and patterns at intervals and between these very goliath social economic datasets. These knowledge hold the potential to permit call manufacturers to trace development progress, improve social protection, and perceive wherever existing policies and programmers need adjustment. This paper presents the novel challenges and opportunities related to massive knowledge necessitate rethinking several aspects of those knowledge management stages, whereas holding alternative captivating aspects.

Nasser T and Tariq RS et al [4] "Big knowledge Challenges", the management cluster includes the legal and good problems associated with accessing knowledge bedded design reference referred to as "big knowledge technology stack" will be conferred as theoretical resolution framework for the challenges of the massive knowledge. Every layer can give the technologies needed to beat very surprising challenge however mutually of these layers give the entire resolution. Continues evolution of technology necessitate developing new massive knowledge investigation to delve a considerable measure of deeper into the data searching for a ton of valuable bits of knowledge and cathartic new massive knowledge version a couple of.0.

Yuri Demchenko, Zhiming Zhao et al [5] "Addressing massive knowledge Challenges for Scientific knowledge Infrastructure" This paper discusses the challenges that are compulsory by massive knowledge Science on the fashionable and future Scientific knowledge Infrastructure (SDI). The paper refers to very surprising scientific communities to outline necessities on knowledge management, access management and security. The paper introduces the Scientific knowledge Lifecycle Management (SDLM) model that has all the primary stages and reflects specifics in knowledge management in trendy Science. The paper proposes the SDI generic design model that gives a reason for building functional knowledge or project centrical SDI exploitation trendy technologies and best practices. The paper explains however the planned models SDLM and SDI are often normally enforced exploitation trendy cloud fundamentally based infrastructure services provisioning model.

3. CHALLENGES

The challenges can be summarized as follows:

*   **Volume**: The data is being generated at a very high speed. It was in peta bytes in 2000 and it will be in Zeta bytes by 2020.

*   **Variety:** Data available is in many forms. It is hardly available in structured form. It is available as raw, semi-structured, unstructured along with data from web pages, search indexes, images and videos etc.

*   **Velocity:** The flow of data coming in and out of the system is very high. In general, there is no particular technology is available to handle the flow of information.

*   **Quality:** The data available must be relevant to the context of the problem and determining the quality of the data available for a particular problem is a major challenge.

*   **Privacy and Security:** While finding out useful information from data, data scientists usually ignores the privacy and security of the persons concerned. Care must be taken so that people should not compromise with their privacy.

*   **Scalability:** It is the ability to process huge amount of data in a single application and it is challenging to have unlimited scalability.

*   **Veracity:** It means uncertainties, untruths and missing values in the data. It measures the accuracy of the data so that it can be used for data analysis. It is the most important challenge for the data scientists.

*   **Volatility:** It means for how much time data will be valid i.e. for how much time it

Sanjeev Kumar*

has to be preserved in databases and will be valuable.

- **Discovery:** It means how to find out high quality data from huge amount of raw data available.

- **Dogmatism**: It means after extracting useful information from the data, the data scientists/researchers should apply common sense, consult domain experts to verify the validity of the output.

- **Process Challenges:** It contains all the challenges that start from capturing stage and ends when output is presented to the clients. It includes:

  ○ Data Acquisition and recording challenges

  ○ Information extraction and cleaning challenges

  ○ Data Integration challenges

  ○ Query processing and analysis challenges

4. CONCLUSION

In recent years, big data has been generated at a dramatic pace. Analyzing this knowledge is difficult for a general man. To this finish during this paper, we tend to survey the assorted analysis problems, challenges of big data. From this survey, it's understood that each huge knowledge platform has its individual focus. A number of them square measure designed for batch processing whereas some square measure sensible at period analytic. Each big knowledge platform conjointly has specific practicality. Different techniques used for the analysis embody applied math analysis, machine learning, data processing, intelligent analysis, cloud computing, quantum computing, and knowledge stream process. The future researchers can pay a lot of attention to the techniques to resolve issues of massive knowledge effectively and efficiently.

REFRENCES

[1] Shashi Shekhar (2012). "Spatial big-data challenges intersecting mobility and cloud computing", Proceedings of the 11th ACM International Workshop on Data Engineering for Wireless and Mobile Access - In Conjunction with ACM SIGMOD / PODS 2012 (pp. 1-6).

[2] Alexandru Adrian (2013). "Big Data Challenges", Database Systems Journal vol. I, no 3/2013.

[3] Nasser T. and Tariq R.S. (2015). "Big Data Challenges", Nasser and Tariq, J Comput Eng Inf Technol 2015, 4:3.

[4] Yuri Demchenko, Zhiming Zhao (2016). "Addressing Big Data Challenges for Scientific Data Infrastructure", 2016, pp. 120-130.

[5] R. J. Dodd (2016). Monthly Notices of the Royal Astronomical Society, pp. 959 – 972.

[6] Vid Podpečan ; Monika Zemenova ; Nada Lavrač "Orange4WS Environment for Service-Oriented Data Mining" The Computer Journal, 2012, pp. 82 – 98.

[7] Zhenyu Zhou ; Houjian Yu ; Chen Xu ; Yan Zhang ; Shahid Mumtaz ; Jonathan Rodriguez (2018). "Dependable Content Distribution in D2D-Based Cooperative Vehicular Networks: A Big Data-Integrated Coalition Game Approach" IEEE Transactions on Intelligent Transportation Systems, pp. 1-12.

[8] Ivano Notarnicola ; Ying Sun ; Gesualdo Scutari ; Giuseppe Notarstefano (2017). "Distributed big-data optimization via block-iterative convexification and averaging", Decision and Control (CDC), 2017.

[9] Mostafa Rahmani ; George Atia (2017). "Robust and Scalable Column/Row Sampling from Corrupted Big Data" Computer Vision Workshop (ICCVW), pp. 112-125.

[10] K. Kambatla, G. Kollias, V. Kumar and A. Gram (2014). Trends in big data analytics, Journal of Parallel and Distributed Computing, 74(7), pp. 2561-2573.

[11] R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese (2013). A look at challenges and opportunities of big data analytics in healthcare, IEEE International Conference on Big Data, pp. 17-22.

[12] Ammu, Nrusimham, and Mohd Irfanuddin (2013). "Big Data Challenges." International Journal of Advanced Trends in Computer Science and Engineering 2.1: pp. 613-615.

[13] Che, Dunren, Mejdl Safran, and Zhiyong Peng (2013). "From big data to big data mining: challenges, issues, and opportunities." International Conference on Database Systems for

Advanced Applications. Springer Berlin Heidelberg.

[14]   Judith H., Alan N., Fern H., Marcia K. (2013). Big Data for Dummies, John Wiley &Sons, Inc., New Jersey, USA.

[15]   Ohlhorst F.J. (2012). Big Data Analytics: Turning Big Data into Big Money, John Wiley & Sons, 176.

[16]   O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis and K. Taha (2015). Efficient machine learning for big data: A review, Big Data Research, 2(3), pp. 87-93.

[17]   P. Singh and B. Suri (2014). Quality assessment of data using statistical and machine learning methods. L. C. Jain, H. S. Behera, J. K.Mandal and D. P. Mohapatra (eds.), Computational Intelligence in Data Mining, 2, pp. 89-97.

Corresponding Author

Sanjeev Kumar*

Assistant Professor, Department of Computer Science, DAV College, Abohar, Punjab, India

E-Mail – gumber_sanjeev@yahoo.com