CHALLENGES AND ISSUES IN BIG DATA

M.Meena Krithika[1], K.Kannika Parameswari[2], K. Gayathri[3]
Assistant Professor, Department of CS, NGMC, Pollachi,
mmeenakrithika@gmail.com[1],kannika.krish@gmail.com[2],gayathri.it91@gmail.com[3]

ABSTRACT:

The espouse of data-driven determination making is being conceived broadly, and there is enlarging excitement for the idea of "Big Data". While the espouse of Big Data is genuine, there is parallel gap within its potentiality and its attainment.

Heterogeneity, complexity, timeliness, progress, and recession problems with Big Data prevent proceeding at all stages of the pipeline that can produce value from data. The obstacle starts instantly during data appropriation, when the data obligate us to make determination currently in an ad hoc manner. Most of the data is not related in structured format.

The assessment of data blow up when it can be combined with another data, thus data adherence is a greater creator of evaluation. Since many data is instantaneously originated in digital format today and we have the scope and denounce to persuade the creation to propitiate subsequently and to automatically link previously completed data. Data analysis, association, gathering and modeling are other basic challenges. Data analysis is a open bottleneck in several applications, both due to requirement of scalability of the domestic algorithms and due to the complication of the data that needs to be resolved.

The many unique challenges and opportunities comparable with Big Data involve rethinking several outlooks of these data management stages, while maintaining other beneficial aspects. We conclude that particular investment in Big Data will lead to a new wave of fundamental technological growth that will be embodied in the next origination of Big data management and analysis platform.

We conclude that these research issues are not only instant, but also have the ability to create large economic value in the economy for years to come. However, they are also difficult and requiring us to rethink data analysis systems in crucial ways. A dominant investment in Big data, efficiently directed, can flow not only in major scientific advances, but also lay the footing for the next origination of advances in medicine, science, and business.

Keywords: *Big Data, Challenges in Big Data, Issues in Big Data.*

INTRODUCTION:

With the accession of the digital age, the effect of data being originated, shared and stored has been on the rebel. Data is the primary element of communication and accordance in internet and all the applications that are produced on this platform. In a large range of application areas, data is being gathered at phenomenal scale. Determinations that previously were depended on guesswork, or on studied models of presence, can now be carried based on the data itself. Such Big Data analysis presently forced nearly every outlook of our modern society, including retail, financial services, manufacturing, life science, mobile services, and physical sciences. This data has enormous ability, ever-increasing structure, insecurity and venture and irrelevance. The advantages and limitations of accessing this data are disputable in view of the fact that this analysis may involve access and analysis of social media communications, medical records, government records, financial data, and genetic outcome.

Big Data has the ability to remodel not just an exploration, but also education. While the potential advantages of Big Data are real and vital, and some initial achievements have already been concluded, that endure many technical challenges that must be discovered to fully accomplish this potential. The analysis of Big Data comprises of multiple different phases and each of which introduce challenges. In this paper, we consider the challenges and issues of the Big Data along with the future research work.

CHALLENGES AND ISSUES IN BIG DATA:

In order to move beyond the existing techniques and strategies used for machine learning and data analytics, some challenges need to be overcome. In order to choose a sufficient method or design, a solid scientific groundwork needs to be developed. For proper implementation of devised solutions, appropriate development skills and technologies platforms must be identified and developed. We now turn to some common challenges underlying in Big Data.

When man utilizes information, a remarkable act of heterogeneity is easily tolerated. In fact, the intensity and richness of fundamental language can assist valuable intensity. However, machine analysis algorithms insist homogeneous data, and cannot conceive intensity. In elevation, data must be thoughtfully arranged as a first step in (or preceding to) data analysis. Consider, for example, a patient who has multiple medical operations at a hospital. We could create one record per medical operation or laboratory test, one record for the complete hospital stay, or one record for all lifetime hospital connection of this patient. With any other than the first design, the number of medical operations and lab tests per record will be differ for each patient. However, computer system work most sensibly if they can store aggregated items that are all matching in size and structure. Suitable Illustration, access, and inquiry of semi structured data needs further work.

Even later data cleaning and error alteration, some imperfectness and some errors in data are possible to remain. This imperfection and these errors must be arranged during data analysis. Doing this accurately is a challenge. Advanced work on using probabilistic data proposes one way to make improvement.

Handling large and fast increasing volumes of data has been a challenging outcome for many decades. In the past, this challenge was moderated by processors getting fixed, to provide us with the resources essential to cope with enlarging volumes of data. But, there is an original shift underway now. Data volume is measuring faster than compute resources. Unfortunately, similar data processing methods that were utilized in the past for computing data across nodes don't immediately applied for intra-node parallelism, since the structure looks very different. These phenomenal changes direct us to rethink how we plan, build and engage data processing fundamental.

The recession of data is one huger problem, and one that enlarge in the context of Big Data. For maintain health records, there are severe laws governing what can and cannot be done. For other data, arrangements are less potent. However, there is eminent public fear concerning the unsuitable use of personal data, especially through linking of data from many sources. Handling privacy is effectively a technical and sociological problem, which must be noted from both standpoints to realize the assurance of big data. The research regulations are not restricted to the above-mentioned points. The main aim is to convert the cloud from being a data organization and infrastructure basics to a scalable data analytics fundamental.

CONCLUSION:

We have entered a period of Big Data. Through greater analysis of the huge volumes of data that are seemly available, there is the possible for making secure advances in many external disciplines and developing the profitability and success of many businesses. However, many technical contests summarized in this paper must be described before this potential can be accomplished fully. This challenge includes not just the distinct issues of scale, but also heterogeneity and lack of structure. These challenges are ordinary across a huge variance of application areas, and therefore not cost sufficient to describe in the context of one area alone. Furthermore, these challenges should need alternate solutions, and will not be described naturally by the next formation of industrial products. We must maintain and nourish fundamental research around addressing these types of technical challenges if we are to accomplish the intended advantages of Big Data. Research analyzed in the area of Big Data analytics aim to constitute an energetic and efficient system that describes the identified challenges and concerns.

REFERENCES:

[1] Agarwal, D., Das, S. and Abbadi, A. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. ACM 978-1-4503-0528-0/11/0003. Retrieved from: http://www.edbt.org/Proceedings/2011-Uppsala/papers/edbt/a50-agrawal.pdf

[2] Amazon Kinesis. (n.d.). Developer Resources. Retrieved from: http://aws.amazon.com/kinesis/developer-resources/

[3] Apache S4. (n.d.). Distributed Stream Computing Platform. Retrieved from: http://incubator.apache.org/s4/

[4] Assuncao, M. D., Calheiros, R. N., Bianchi, S. and Netto, M. A. S. (2015). Big

Data Computing and Clouds: Trends and Future Directions. J. Parallel Distrib. Computing, 79-80 (2015) 3-15. Retrieved from: http://www.buyya.com/papers/BDC-Trends-JPDC.pdf

[5] Borthakur, D., Gray, J., Sarma, J. S., Muthukkaruppan, K., Spiegelberg, N., Kuang, H., Ranganathan, K., Molkov, D., Menon, A., Rash, S., Schmidt, R. and Aiyer, A. (2011). Apache Hadoop Goes Real-time at Facebook, in: Proceedings of the ACM SIGMOD

International Conference on Management of Data (SIGMOD 2011), ACM, New York, USA, 2011, pp. 1071–1080. Retrieved from: http://cloud.pubs.dbs.uni-leipzig.de/sites/cloud.pubs.dbs.uni-leipzig.de/files/RealtimeHadoopSigmod2011.pdf

[6] Calheiros, R. N., Vecchiola, C., Karunamoorthy, D. and Buyya, R. (2012) The Aneka platform and QoS-driven resource provisioning for elastic applications on hybrid Clouds, Future Gener. Comput. Syst. 28 (6) (2012) 861–870. Retrieved from: http://www.buyya.com/papers/Aneka-QoS-ResourceProvisioning-FGCS.pdf

[7] Chen, Q., Hsu, M. and Zeller, H. (2011). Experience in Continuous analytics as a Service (CaaaS), in: Proceedings of the 14th International Conference on Extending Database Technology, ACM, New York, USA, 2011, pp. 509–514. Retrieved from: http://www.edbt.org/Proceedings/2011-Uppsala/papers/edbt/a46-chen.pdf

[8] Chen, H., Chiang, R. H. L. and Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly. Special Issue: Business Intelligence Research. Retrieved from: http://ai.arizona.edu/mis510/other/MISQ%20BI%20Special%20Issue%20Introduction%20Chen-Chiang-Storey%20December%202012.pdf

[9] Dean, J. and Ghemawat, S. (2004). OSDI 2004. Retrieved from: http://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf

[10] Demirkan, H. and Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. Decision Support Systems 55 (2013) 412-421. Retrieved from: http://www.crisismanagement.com.cn/templates/blue/down_list/llzt_dsj/Leveraging%20the%20capabilities%20of%20service-oriented%20decision%20support%20systems%20Putting%20analytics%20and%20big%20data%20in%20cloud.pdf

[11] GigaSpaces. (2012). Big Data Survey. Retrieved from: http://www.gigaspaces.com/sites/default/files/product/BigDataSurvey_Report.pdf

[12] Google Cloud Platform. (n.d.). Big Query. Retrieved from: https://cloud.google.com/bigquery/

[13] Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F.B. and Babu, S. (2011). Starfish: A Self-tuning System for Big Data Analytics, in: Proceedings of the 5th Biennial

Conference on Innovative Data Systems Research (CIDR 2011), 2011, pp. 261–272. Retrieved from: https://www.cs.duke.edu/starfish/files/cidr11_starfish.pdf

[14] IBM InfoSphere Streams. (n.d.). InfoSphere Streams. Retrieved from: http://www.ibm.com/software/products/en/infosphere-streams.

[15] Kim, H., Abdelbaky, M. and Parashar, M. (2009). CometPortal: A Portal for Online Risk Analytics Using CometCloud. 17th International Conference on Computer Theory and Applications (ICCTA2009). Retrieved from: http://nsfcac.rutgers.edu/CometCloud/sites/nsfcac.rutgers.edu.CometCloud/files/pub/ICCTA2009.pdf

[16] Lee, K. H., Lee, Y. J., Choi, H., Chung, Y.D. and Moon, B. (2011). Parallel Data Processing with MapReduce: A Survey, SIGMOD Record 40 (4) (2011) 11–20. Retrieved from: http://www.cs.arizona.edu/~bkmoon/papers/sigmodrec11.pdf

[17] Li, X., Calheiros, R. N., Lu, S., Wang, L., Palit, H., Zheng, Q. and Buyya, R. (2012). Design and Development of an Adaptive Workflow-Enabled Spatial-Temporal Analytics Framework, in: of Computer Science and its Applications. Proceedings of the IEEE 18th International Conference on Parallel and Distributed Systems (ICPADS 2012), IEEE Computer Society, Singapore, 2012, pp. 862–867. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.5453&rep=rep1&type=pdf

[18] Manekar, A. and Pradeepini, G. (2015). A Review on Cloud-based Big Data Analytics. ICSES Journal on Computer Networks and Communication (IJCNC), May 2015, Vol. 1, No, 1. Retrieved from: http://www.i-cses.com/ijcnc/Archive/V1N1/IJCNC-V1N1-P0001.pdf

[19] Neaga, I. and Hao, Y. (2014). A Holistic Analysis of Cloud Based Big Data Mining. International Journal of Knowledge, Innovation and Entrepreneurship. Volume 2 No. 2, 2014, pp. 56—64. Retrieved from: http://ijkie.org/IJKIE_December2014_IRINA&HAO.pdf

[20] NESSI. (2012). Big Data: A New World of Opportunities. Retrieved from: http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf