



*Journal of Advances and
Scholarly Researches in
Allied Education*

*Vol. X, Issue No. XX,
Oct-2015, ISSN 2230-7540*

**AN ANALYSIS UPON VARIOUS MODELING AND
EVOLUTIONARY OPTIMIZATION OF BIG DATA
ANALYTICS**

AN
INTERNATIONALLY
INDEXED PEER
REVIEWED &
REFEREED JOURNAL

An Analysis upon Various Modeling and Evolutionary Optimization of Big Data Analytics

Vakala Ramakrishna Sumant¹ Dr. Bechoo Lal²

¹Pebble Hills University, USA

²Assistant Professor, Western College University of Mumbai

Abstract – *In the recent times the amount of data are generated and stored by various industries are rapidly increasing on the internet thus data scientists are facing a lot of challenges for maintaining a huge amount of data as the fast growing industries require the significant information for enhancing the business and for predictive analysis of the information. This paper focuses on the various states of the studies towards Big Data analytic techniques and gives a better comparative analysis of various applications proposed till date. Inference has been done for evaluating the performance efficiency, limitations and the advantages of the different types of existing Big Data Analytic techniques. The main objective of the proposed study is to provide a better and significant research perspective and an overview of data analysis techniques which are referred to the papers found on the web which will be quite helpful for the future research prospective of this domain.*

Keywords: *Big Data, Hadoop, Big Data analytics etc.*

-----X-----

INTRODUCTION

We are living in the era of Big Data. Today a vast amount of data is generating everywhere due to advances in the Internet and communication technologies and the interests of people using smartphones, social media, Internet of Things, sensor devices, online services and many more (Russom, 2011). Similarly, in improvements in data applications and wide distribution of software, several government and commercial organizations such as financial institutions, healthcare organization, education and research department, energy sectors, retail sectors, life sciences and environmental departments are all producing a large amount of data every day (Shadi, *et. al.*, 2008). For examples, International Data Corporation (IDC) reported that 2.8 ZB (zettabytes) data of universe were stored in the year of 2012 and this will reach up to 40 ZB by 2020. Similarly Facebook processes around 500 TB (terabytes) data per day and Twitter generates 8 TB data every day. The huge datasets not only include structured form of data but more than 75% of the dataset includes raw, semi-structured and unstructured form of data. This massive amount of data with different formats can be considered as Big Data (Tech America, 2012).

technologies and the interests of people using smartphones, social media, Internet of Things, sensor devices, online services and many more. Similarly, in improvements in data applications and wide distribution of software, several government and commercial organizations such as financial institutions, healthcare organization, education and research department, energy sectors, retail sectors, life sciences and environmental departments are all producing a large amount of data every day (Talia, 2013. Kwan, Muelder, 2013. Cevher, *et. al.*, 2014). The huge datasets not only include structured form of data but more than 75% of the dataset includes raw, semi-structured and unstructured form of data. This massive amount of data with different formats can be considered as Big Data.

REVIEW OF LITERATURE:

Today a vast amount of data is generating everywhere due to advances in the Internet and communication

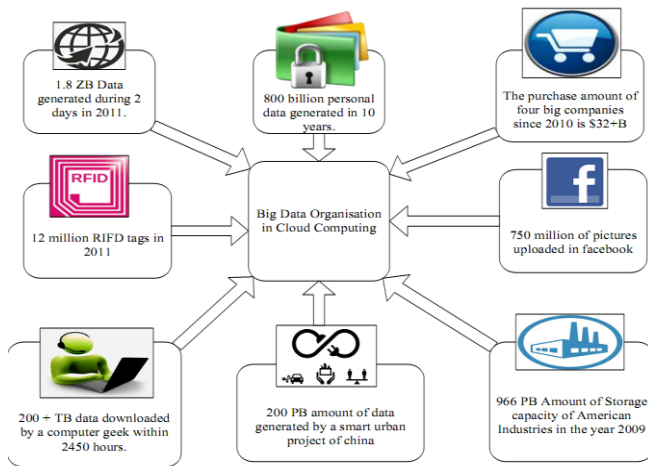


Fig 1: The Uninterruptedly Increasing Big Data

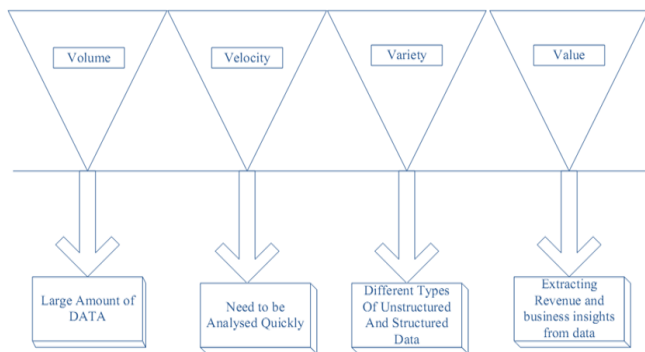


Fig 2: Characterization and features of Big Data

1. Connectivity between Cloud Computing and Big Data: The concepts of Cloud computing and big data are co-related to each other. Big data is considered as an object of the computation oriented operations which increase the stress over various storage capacity of the cloud computing system. The main objective of the cloud computing is to handle a big amount of data applications with the efficient, fine-grained and low computational complexity with sufficient storage capacity and processing resources (Hu, *et. al.*, 2014). The development of the cloud computing infrastructure provides an ease of storage management, computing and processing of huge amount of data. It can be said from a different perspective that big data also accelerates the growth of cloud computing infrastructure (Talia, 2013). The Following figure provides an overview of cloud computing management.

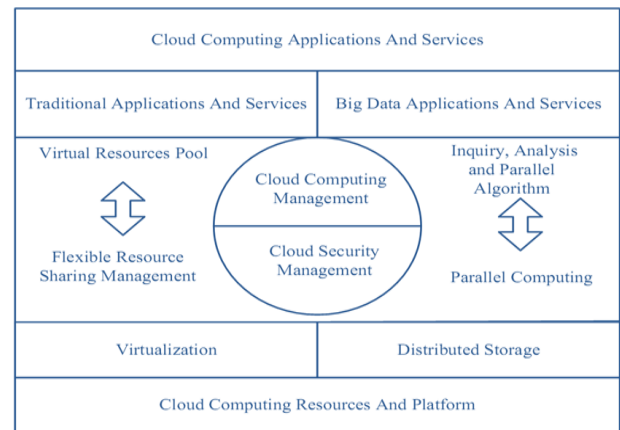


Fig 3: Cloud Big Data Relationship

2. Big Data Storage: As the growth of Data is happening explosively thus it requires efficient storage management system. This section discusses briefly the storage of Big Data. The functionality of big Data Storage includes the management of large scale and unstructured data sets with the ease of reliability and availability of the data accessing (Wu, *et. al.*, 2014). Various studies are concerned about discussing the essential issues associated with the massive storage system, dispersed storing organizations, and big data storage systems.

3. Connectivity between Hadoop and Big Data: Presently Hadoop has become an indispensable environment for application development associated with Big Data for example junk mail sifting, web searching, click stream analysis etc. In the recent times, most of the researches are more focused on Big Data analytics based on Hadoop technology. Some of the representative discussion shows that in June 2012 Hadoop has been run by yahoo in different servers with four data centers to for supporting various products and services.

4. Big Data Analysis: The data analysis concept of Big Data gives various analytical methods which can be applicable for analyzing traditional datasets which includes various analytical architecture, software requirement for exploration of big data (Tan, 2013). Data investigation is one of the most essential stages of the big data value chain where the main objective is to extract the meaningful information and providing suggestions and decisions. Different types of possible and gravitational values can be produced through the several stages of analysis in different fields. Data Analysis is considered to be a very broad area where the environment is so complex and associated with the use of various complex methods, architectures, and tools.

CONCLUSION:

As we have entered an era of Big Data which is the next frontier for innovation, competition and productivity, a new wave of scientific revolution is about to begin. Fortunately, we will witness the coming

technological leapfrogging. In this survey study, we give a brief overview on Big Data problems, including Big Data opportunities and challenges, current techniques and technologies. We also propose several potential techniques to solve the problem, including cloud computing, quantum computing and biological computing. Although those technologies are still under development, we have confidence that in the coming future we will receive several great breakthroughs in those areas. Undoubtedly, today and future's Big Data problems will benefit from those progresses. There is no doubt that Big Data analytics is still in the initial stage of development, since existing Big Data techniques and tools are very limited to solve the real Big Data problems completely, in which some of them even cannot be viewed as Big Data tools in the true sense.

REFERENCES:

- Russom, P. (2011).: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40.
- Shadi Ibrahim, Hai Jin, Lu Lu (2008). "Handling Partitioning Skew in MapReduce using LEEN" ACM 51, pp. 107–113
- Tech America (2012) : Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: TechAmerica Reports, pp. 1–40.
- Talia, D. (2013). Clouds for Scalable Big Data Analytics, Computer, Vol.46, No.5, pp.98-101.
- Kwan-Liu Ma; Muelder, C.W. (2013). Large-Scale Graph Visualization and Analytics. Computer , vol. 46, no. 7, pp. 39,46.
- Cevher, V.; Becker, S.; Schmidt, M. (2014).Convex Optimization for Big Data: Scalable, randomized, and parallel algorithms for big data analytics. Signal Processing Magazine, IEEE , Vol.31, No.5, pp. 32-43.
- Hu, H; Wen., Y; Chua, Tat-Seng., Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. Access, IEEE , Vol.2, pp. 652-687.
- Wu, L., Barker, R.J., Kim, M.A., Ross, K.A. (2014). Hardware Partitioning for Big Data Analytics. Micro, IEEE, Vol.34, No.3, pp. 109-119.
- Tan, W., Blake, M.B., Saleh, I., Dustdar, S. (2013). Social-Network-Sourced Big Data Analytics," Internet Computing, IEEE, Vol.17, No.5, pp. 62-69.