

State of the Art of Big Data Analytics: A Survey

Rajeshwari.D
Assistant Professor
NIE Institute of Technology
Mysore, Karnataka, Bangalore

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 art 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, Cloud Computing, Hadoop, Big Data analytics.

1. INTRODUCTION

Since past few years the huge amount of data has expanded continuously in various arenas thus Big Data Analytics has become one of the most discussed topics among researchers for further research direction. The term Big Data is used to refer the enormous amount of datasets. Compared to the traditional datasets big datasets comprises a set of unstructured data which requires more significant real-time investigation. The concept of big data helps us to understand the abstraction and in-depth understanding behind various hidden values. Recently industries are more interested in big data Analysis and various government agencies are declared many plans for increasing the research towards big data analytics. Now a days the big data associated with various industries are growing expeditiously hence Google which is one of the leading software giants in the market processes hundreds of Petabyte amount of data, Face book and a Chinese organization named Baidu generates 10 PB amount of data per month respectively. The rapid development of the combination of cloud computing phase and the Internet of Things (IoT) endorses a sharp growth of data. Big data oriented cloud computing delivers an ease of storage management as various applications and data resources are spread among the users of the worldwide in a distributed manner [1]. The model of IoT (Internet of Things) includes a huge amount of data collection by various sensor nodes and transmission of the data over the cloud network for storing and further processing. The most essential benefits of cloud

computing infrastructure includes various resources such as availability , elasticity and cost reduction in various amount of data in a process where one has to pay before using a particular service. Cloud infrastructure also provides some benefits to the organizations where they do not have to pay the maintenance cost for the IT infrastructure. In a cloud environment data analytics and analysis should be very strong as there are many challenges that can be arises. This study highlights various approaches related to Big Data analytics and the development of Analytics solution for optimizing various issues of cloud.

It is observed that now a days various organizations are consuming enormous amount of essential information which can be useful in various fields such as monitoring of an objects activity, sensor deployment , tracking of data etc. In age of Facebook , Twitter and other social medias many users express their daily activities , and various kinds of events they attend as well as many pictures , videos for recording such things in a cloud platform. A severe flood associated with data is termed as Big Data, which yields to the challenging situation on the present infrastructure of Data Storage management and the Statistical estimation of data. Competitive situations arises when an organization wants to explore its data from it personal websites for analyzing the customer's feedback, customized services towards a product. As a result the various decisions makers would convey their conclusions grounded on the analysis of extracted data or those data which carry some value or weightage. The Knowledge Discovery Process which is known as (KDD) is used for extracting valuable information from an unstructured data sets after completing detailed analysis and interpretations. Data mining is also mapped between various unrelated attributes of datasets which can be obtained from machine learning, database systems and statistics. There various analytics procedures are present which includes statistical and quantitative analysis etc. . The development of cloud computing integrated the flexibility of IT industries by providing a huge amount of resources and services. The concept of Cloud infrastructure says that an organization has to pay for using operational services where different types of resources are needed for running some particular applications. It can be seen that some specific technologies and implementation are required for the cloud for providing infrastructure, platform and software services.. Fig 1. Shows the uninterruptedly increasing big data which have been generated since many years [2]. The fig also describes that how organizational data, social networks data are enormously increasing now a days for an example 750 million of pictures have been uploaded on Facebook servers till now.

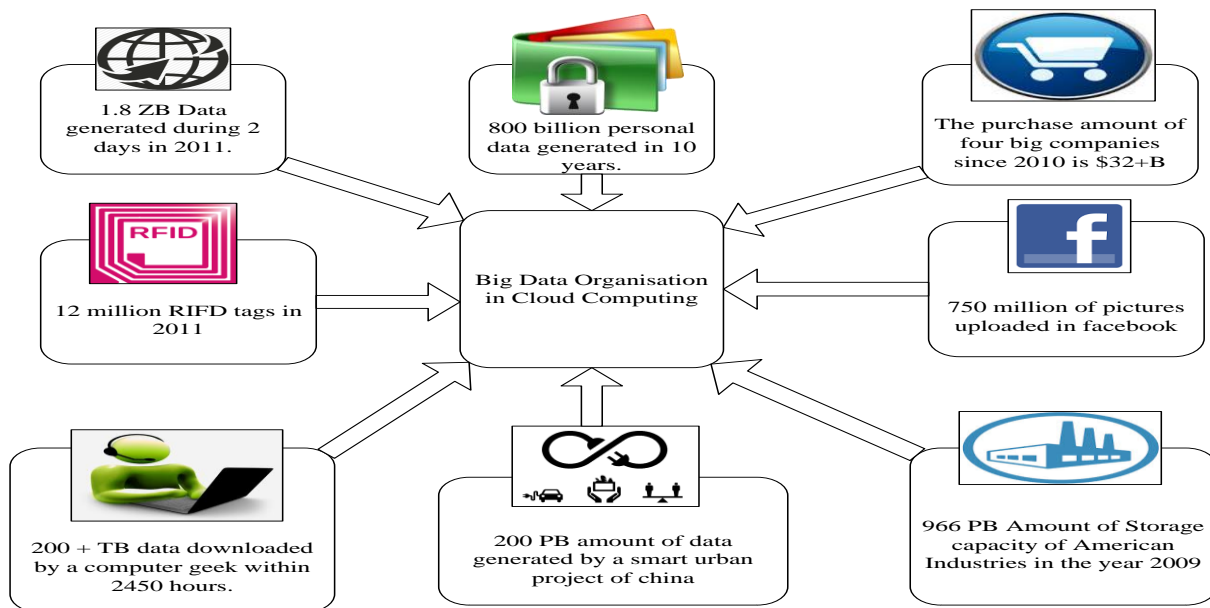


Fig 1: The Uninterruptedly Increasing Big Data

It has been observed by the McKinsey & company that how big data generates significant values after in-depth research which was experimented in US health care service, EU public sector administration, the US retail, the global manufacturing, and global personal location data. The proposed research in five core industries shows the global economy. The annual

report of McKinsey also presents an occupied production to the commercial function; it also measures the productivity, competitiveness and the growth of various industries and public sectors and helps consumers to make profit efficiently with the quality of services.

Big Data: The Four Vs

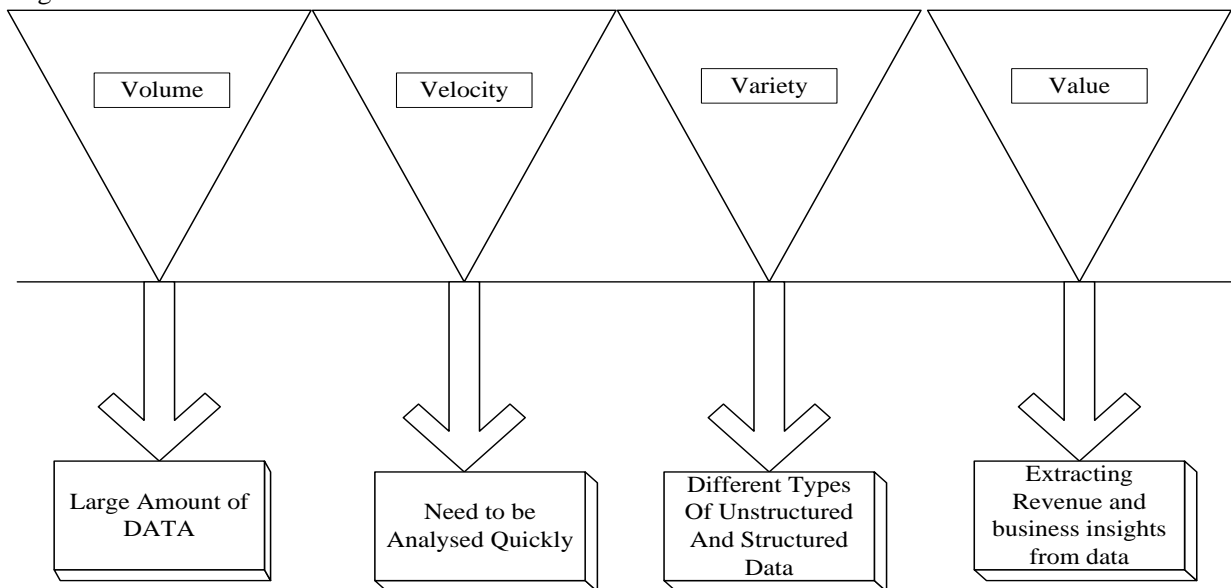


Fig 2: Characterization and features of Big Data

1.1 The Expansion of Big Data

The perception of database contraction arises in the time of 1970s where this technology has been created for maintaining and analyzing data. As the amount of data increasing day by day thus the storage management and computing capability of a single mainframe system have become very insufficient. In the era of 1980s share nothing which is a parallel database system has been introduced for overcoming the issues related to the data volume increasing. The system architecture of share nothing was defined in the basis of cluster principles,

where every considered machine has their own processor, storage and disk space management system [3]. Teradata system is considered to be a first most effective commercial distributed system though its database gained some popularity lately. On June 2, 1986 the study of Walter T has been considered as milestone event where the Teradata concept first provides the first parallel distributed system mechanism with the storage space capability of 1TB to a large scale retail organization named Kmart for in North America. Therefore various big giants like Google, Amazon, and Face book etc.

invested in this area and started doing various projects of big data analytics.

1.2 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. 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 [4]. The Following figure provides an overview of cloud computing management.

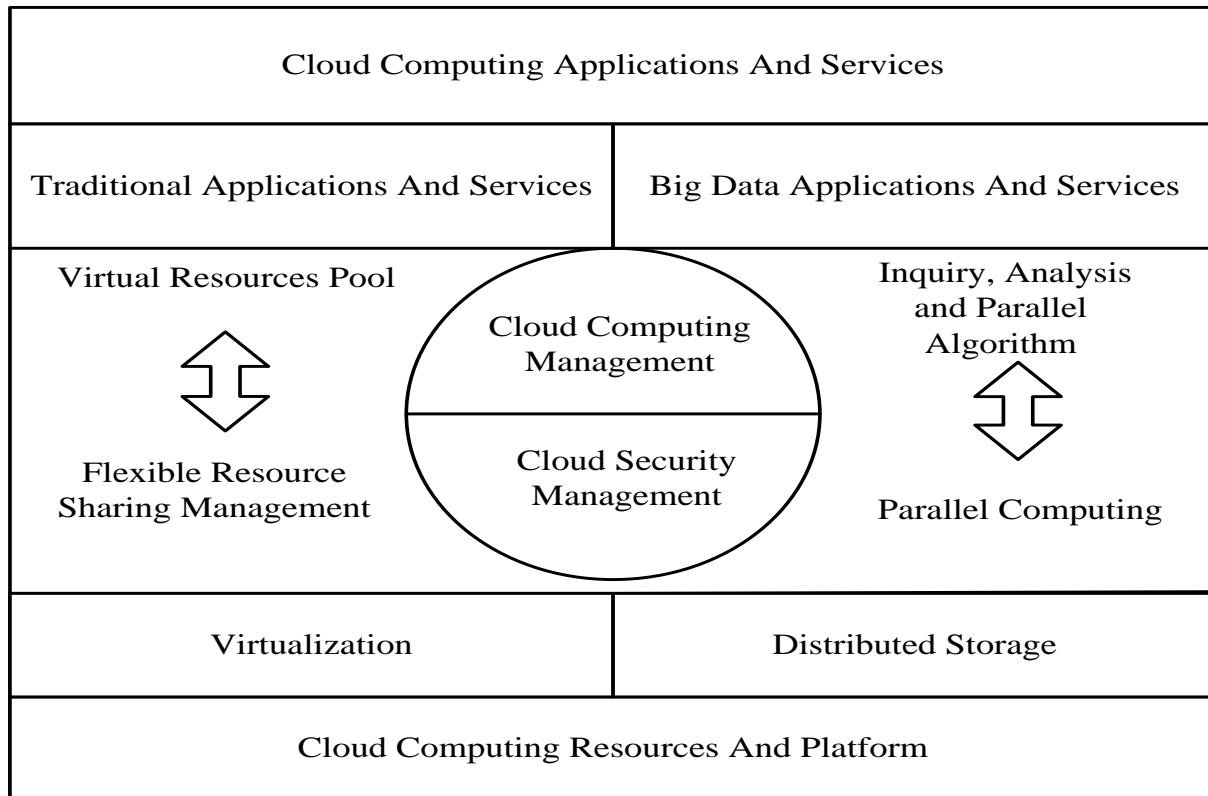


Fig 3: Cloud Big Data Relationship

1.3 Connectivity between IoT and Big Data

In the real world scenario of IoT which is known as Internet of things involves an enormous amount of sensor nodes which can be entrenched into various devices and machines for accomplishing a particular job. In order to collect various kind of information sensor nodes are deployed in many hostile, inaccessible and dynamic environments. The nodes functionality is to sense the environment to gather the various type of information such as environment data, geographical information etc. There is some difference of characteristics in between IoT generated big Data and normal Big Data as in different paradigms different types of data are gathered. The most classical characteristics of big data include heterogeneity, variety, unstructured features, noise and high redundancy. It is also observed that the IoT generated data are not that much essential part of big data analytics and It is also analysed according to a report of HP that by 2030 the number of sensor nodes will reach up to 1 trillion and that time it will become a very difficult scenario to manage all the data generated by IoT in Big Data environment. A report published by Intel also highlights the Big Data and IoT related features which are 1. Plentiful terminals generating masses data 2. Data which are produced by IoT, which is usually

semi-structured and unstructured 3. Data of IoT have no use without examining the essential features [5].

1.4 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. They recent survey highlights that the biggest Hadoop cluster consists of 4000 nodes but in future the number of nodes will be extended to 10000 with the release of Hadoop 2.0. A statistical estimation of Face book says that its own Hadoop clusters can process 100 PB data. Hadoop also used by some well-known organization for distributed cloud computing thus many companies provide various Hadoop commercial executions which include Cloud era, IBM, Map R and Oracle [6].

1.5 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. Various studies are concerned about discussing the essential issues associated with the massive storage system, dispersed storing organizations, and big data storage systems. The big data storage paradigm is designed for providing data storage service with reliability and authenticity. On the other hand, the storage amount should also provide the flexibility for data accessing, query and analysis of large amount of data. As an auxiliary equipment of server data storage devices are used for managing and analyzing the unstructured datasets with structured RDBMS. The rapid growth of data highly requires some efficient data storage device for managing a large amount of data. But many companies own their big cloud storage capacity but which obtains cost effectiveness and it is competitive in nature thus there is a compelling need of research direction towards data storage [7].

1.6 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. 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.

1.7 Big Data Applications

Recently Big Data Analytics has emerged as an advance analytical Technology where large-scale and multifaceted data are processed under specific analytical methods, considered as a part of it, thus various data-driven applications are combined in the past decade. As an example it can be said that in early 1990s BT has been considered as prominent technology in the field of business application development but the search engine based massive data mining processing has come into the market in 21st century. Different types of probable and significant applications and their data investigation characteristic are discussed below.

Evolution of Commercial Applications: Initially the business data was generally considered as a structured data, and composed by companies from inheritance systems and then stored in RDBMSs [8].

- **Evolution of Scientific Applications:** A Big Data analytics survey recently introduced by The U.S. National Science Foundation (NSF) for endorsing the competency to extract the significant data and insights from large and complex collections of digital data. For example, in biology, i-Plant applies to the network infrastructure, physical calculating capitals, organization atmosphere, computer-generated mechanism resources, Operational analysis software, and information provision to researchers, academicians and students in inspiring herbal disciplines.

1.8 Limitations in Big Data

The Enormous growth of data brings out various challenges associated with data achievement, Storage management, and analysis. Various limitations of Big Data program are highlighted below.

- **Data Representation:** It is really difficult to characterize various types of datasets where it is found that a certain levels of heterogeneity in type are present such as structure, semantics, organization, granularity, and accessibility. The main objective of the Data representation is to provide meaningful and structured data for analysis and user interpretation.
- **Redundancy Optimization and Data Compression:** It is also observed in the area of Big Data analytics that presence of high level complexities in the datasets makes the exploration of data analysis very challenging. With the Redundancy optimization and data looseness techniques, it is easy to optimize the indirect cost of the entire system without affecting the possible standards of the data.
- **Data Life Cycle Management:** Data life cycle management includes various challenges associated with the slow advances of storage management systems where a lot of processing challenges with current storage system faces scalability disorders.
- **Analytical Mechanism:** In The traditional RDBMS concept designing, scalability and expandability considered as two main issues thus the investigative arrangement of big data executes diverse data within a limited time.
- **Confidentiality:** In the recent times most of the big data service providers cannot able to efficiently maintain and analyze the big datasets as they have very limited computing capability and resources. The big organizations must depend on the other professional or services for analyzing the data which increases the possibility of the vulnerable attack on data and poses safety risks.
- **Energy Management:** the energy depletion issues of mainframe computing systems grab the attention of the researchers from both the economy and environmental perspective as the processing, storage management and transmission of enormous amount of data consumes more energy.

This study is more focused towards the detail analysis of various existing techniques of Big Data Analytics and gives a brief overview of various techniques discussed in the Table I. The proposed study also discusses about the review of literature and the research gap found in the two decade investigation of Big Data. Conclusion and the future work are highlighted in the end of the paper. The main contribution of this paper is the research gap and the further direction of the research [9].

2. RELATED WORK

The following table highlights the comparative analysis of various Techniques associated with the Big Data Analytics. This Brief discussion will be quite helpful for the future researchers for overcoming various limitations associated with Big Data in Cloud Computing.

The Table I discusses about the various existing techniques and their determination in solving many identified constraints

in the area of Cloud Computing and Big Data analytics research environments. Where majority of the techniques works on managing a large amount of datasets and complicated data structures of Big Data. Very less papers have been found which discusses about the Throughput and energy efficiency of Big Data paradigm.

This paper mostly focuses on various technical issues associated with Big Data Analytics of Cloud Computing. In addition the performance analysis of various techniques and the set of research oriented issues are highlighted in section IV which will be beneficial for the current researchers on future direction of Cloud supported Big Data Computing.

Table.1: Existing Techniques Associated with Big Data in Cloud Computing

| Author | Problems Identified | Techniques Applied | Interpretation |
|------------------------|---|--|--|
| Hu et al [10] | Unstructured Big Data Requires More Real-Time Analysis. | A Framework To Deteriorate Big Data Systems | Benchmark Studies And Possible Research Guidelines For Big Data Systems. |
| Slavakis et al [11]. | Managing Large Amount Of Big Data | Forth Encompassing Models | Decentralized And Lively Approximation, Prediction, and Of Network Link in terms of Load Traffic. |
| Srinivashan et al [12] | Large Volumes Of Electronic Data | Two Novel Applications That Leverage Big Data To Detect Fraud, Abuse, Waste, And Errors In Health Insurance Claims | Benchmarks on the basis of Data Assessment And Potential Research Directions For Big Data Systems. |
| Zhang et al [13]. | Multitasking Workloads For Big-Data. | Ordinal Optimization Using Rough Models And Fast Simulation | Experimental Results Show That Our Evolutionary Approach Compared With Existing Methods, |
| Simmhan et al [14] | Dynamic Demand Response | Smart Grid Cyber-Physical Sagile-system | Examines The Role Of Clouds And Their Trade-Offs. |
| Fu et al. [15] | Data Structure Is Increasingly Complicated In Big Data | A Root Cause Detection Framework | The Proposed Approach Has Been Validated By A Sequence Of Simulations |
| Tan et al. [16] | Very Large Datasets | Social Network Paradigm | Personal Ad Hoc Clouds Encompassing Individuals In Social Networks To Address Such Challenges |
| Cevher et al. [17] | Computational, Storage, And Communications Overheads | Approaches which are first imperative and randomized in nature. | The New Big Data Algorithms Are Based On Surprisingly Simple Principles And Attain Staggering Accelerations Even On Classical Problems |
| Slavakis et al. [18] | Big Data Overhead | Sentiment Analysis | Fundamental Contributions In Big Data Theory And Practice |
| Wu et al. [19] | Throughput And Energy Efficiency Of Large-Scale Data Processing | Hardware-Accelerated Range Partitioner (HARP) | The Framework And HARP Provide Improvement Performance |
| Talia et al [20]. | Extracting Knowledge From Big Data | Smart And Scalable Analytics Services, Programming Tools, And Applications | Performance Analysis In The Basis Of Scalable Analytic Services. |
| Otero et al [21]. | Power Consumption Of Big Data | Details Of Big Data Software | Main Problems Come across to the researchers When in Engineering Big Data Proposes Paths For Future Research Direction. |

3. RECENT SURVEY STUDIES

This section gives an overview about various existing review studies found in the website IEEE Xplorer and their inference towards Big Data Analytics. With the search keyword Big Data in Cloud computing the web server returns 8421 conference publications, 799 journals and magazines and 82 e-books as well as 25 open access journals and magazines which talks about Big Data analytics and Big Data Processing. The papers which have been found in the above mentioned

website lies between the ranges of 2006-2015. Fig 4 shows the statistical analysis about the exiting reviews associated with the characteristics of Big Data. It is also found in IEEE Xplorer that most of the existing reviews discuss about the Big Data Value which is one of the most essential features of Big Data Analytics. Table 2 discusses about recent existing reviews oriented issues and the topic which has been focused. After evaluating the existing studies the research gap of the various existing secure routing techniques have discussed in the section IV.

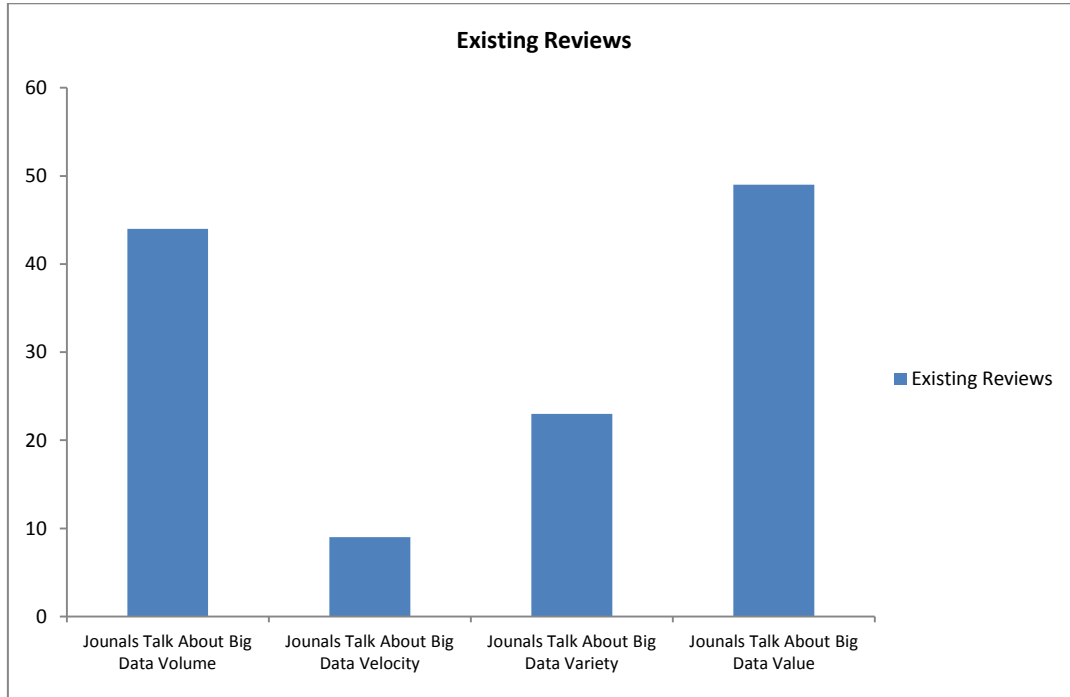


Fig 4 Statistics of Survey on Various Key Features of Big Data in between 2006-2015.

Table 2 Summary of Existing Reviews

| Authors | Year | Topic in focus | Inference |
|-----------------------|------|---|--|
| Ma et al [22] | 2013 | Network Virtualization of Big Data | Pros: Good Theory about domain Cons: No discussion about the prior techniques. |
| Xu et al [23] | 2014 | State of IoT | Pros: Good Theoretical Discussion about prior techniques Cons: No benchmark study followed. |
| Cevher et al [24] | 2014 | Convex optimization algorithms of Big Data | Pros: Discussed various techniques. Cons: Narrowed survey. |
| Srinivasan et al [25] | 2014 | Venture investment in the innovation economy | Pros: Good Theory Cons: No statistical analysis found |
| Hu et al [26] | 2014 | Recent Technological Advancement in Big Data | Pros: Acceptable Discussion on domain. Cons: Narrowed discussion about domain |
| Chelliah et al [27] | 2014 | Transformation and disruption brought by cloud concepts | Pros: Good theoretical overview Cons: Very brief discussion about the prior techniques. |
| Zhou et al [28] | 2014 | Data Analytics of Big Data | Pros: Discussed only few techniques Cons: No statistical review found |

4. RESEARCH GAP

The following highlights the constraints presented in the existing review studies after revising so many papers. The backgrounds and the most recent stages in development of Big Data has been discussed in the proposed theoretical review. It is also found that most of the existing studies related to Big Data, indexed in the Table no.1 are repetitive in

nature. Better Comparative Analysis of the existing studies with respect to the proposed techniques and various performance parameters are significantly missing in the papers [16], [17] and [18]. As the comparative analysis with respect to various performance parameters are very much essential for the for the reader to understand the most efficient techniques till date which assist the reader to continue the

further research for mitigating the various issues related to that domain. It has been also observed that no significant outcomes could be resulted after reviewing the existing studies of Table 1 where the repetitive nature of implementations has been adopted. Most of the recent studies which are presented in Table 2 do not discuss about the prior techniques thus very few benchmarked studies have been found as the majority of the papers highlighted in IEEE Explorer are not benchmarked at all. This situation creates a very much challenging environment for the readers to understand the reliability of the various existing studies towards Big Data and it will be difficult to suggest the further research direction big data. The existing studies proposed by [22], [23], [24] and [27] highlights very less effective performance metrics and narrowed surveys about prior techniques, where Analysis of performance parameters gives a better idea to the reader for understanding the effectiveness of the proposed techniques. As the current research on Big Data is still in an early stage so significant research efforts are needed for analysis and storage management associated with the Big Data.

5. CONCLUSION

Different challenges come out from the applications of Big Data Analytics in Cloud Computing, but this proposed study gives more attention on many aspects which are associated with the decomposition of Big Data System in Cloud Computing. The uniqueness of this paper is that this paper gives an overview of various techniques and highlights most of the significant findings of existing studies which is discussed briefly in Table I. The paper also highlights most of the significant research issues associated with the existing techniques. This survey will be beneficial for the further progress and enhancement of Big Data Analytics in various research perspectives.

6. REFERENCES

- [1] Warneke, D.; Odej K. 2011. Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud. *Parallel and Distributed Systems, IEEE Transactions* , Vol.22, No.6, pp.985-997
- [2] Libin, W., JinLin, W., Peng, S; Jianliang, H. 2008. A Media File Snapshot Technique Based on Embedded System," *Computer Science and Software Engineering, 2008 International Conference*, Vol.4, pp.110-113
- [3] Assunção, Marcos D., Rodrigo N. Calheiros, Silvia Bianchi, Marco AS Netto, and Rajkumar Buyya. "Big Data computing and clouds: Trends and future directions." *Journal of Parallel and Distributed Computing* (2014).
- [4] Warneke, D.; Kao, O. 2011. Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud. *Parallel and Distributed Systems, IEEE Transactions*, Vol.22, No.6, pp.985-997
- [5] Fang, S., Xu, L. D., Zhu, Y., Ahati, J., Pei, H., Yan, J., Liu, Z.2014. An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things. *Industrial Informatics, IEEE Transactions*, Vol.10, No.2, pp.1596-1605
- [6] Hu, H., Wen, Y., Chua, T-S., Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," *Access, IEEE* , Vol.2, No. , pp.652-687, 2014
- [7] Liu, H.2013. Big Data Drives Cloud Adoption in Enterprise. *Internet Computing, IEEE* , Vol.17, No.4, pp.68-71
- [8] Ma, J; Zhang, P., Fu, H-J., Bo, B., Dong, Z-Y.2010. Application of Phasor Measurement Unit on Locating Disturbance Source for Low-Frequency Oscillation. *Smart Grid, IEEE Transactions*, Vol.1, No.3, pp.340,346
- [9] Wang, J., Zhao, P., Hoi, S.C.H.; Jin, R. 2014. Online Feature Selection and Its Applications. *Knowledge and Data Engineering, IEEE Transactions* , Vol.26, No.3, pp.698-710
- [10] Hu, H., Wen, Y., Chua, T-S., Li, X. 2014. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *Access, IEEE*, Vol.2, pp.652-687
- [11] Slavakis, K., Giannakis, G.B., Mateos, G.2014. Modeling and Optimization for Big Data Analytics: (Statistical) learning tools for our era of data deluge. *Signal Processing Magazine, IEEE*, Vol.31, No.5, pp.18-31
- [12] Srinivasan, U., Arunasalam, B.2013. Leveraging Big Data Analytics to Reduce Healthcare Costs. *IT Professional* , Vol.15, No.6, pp.21-28
- [13] Zhang, F., Cao, J., Tan, W., Khan, S.U., Li, K., Zomaya, A.Y.2014. Evolutionary Scheduling of Dynamic Multitasking Workloads for Big-Data Analytics in Elastic Cloud," *Emerging Topics in Computing, IEEE Transactions*, Vol.2, No.3, pp.338-351
- [14] Simmhan, Y.; Aman, S.; Kumbhare, A.; Rongyang L; Stevens, S.; Qunzhi, Z; Prasanna, V.2013. Cloud-Based Software Platform for Big Data Analytics in Smart Grids. *Computing in Science & Engineering*, Vol.15, No.4, pp.38-47
- [15] Chien, C-F., Chuang, S-C.2014. A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics. *Semiconductor Manufacturing, IEEE Transactions*, Vol.27, No.4, pp.475-488
- [16] 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
- [17] 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
- [18] Slavakis, K., Kim, S-J., Mateos, G., Giannakis, G.B.2014. Stochastic Approximation vis-a-vis Online Learning for Big Data Analytics [Lecture Notes]. *Signal Processing Magazine, IEEE* , Vol.31, No.6, pp.124-129
- [19] 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
- [20] Talia, D.2013. Clouds for Scalable Big Data Analytics, *Computer*, Vol.46, No.5, pp.98-101

- [21] Otero, C.E., Peter, A.2015. Research Directions for Engineering Big Data Analytics Software. *Intelligent Systems, IEEE*, Vol.30, No.1, pp.13-19
- [22] Kwan-Liu Ma; Muelder, C.W.2013.Large-Scale Graph Visualization and Analytics. *Computer* , vol.46, no.7, pp.39,46
- [23] Li Da Xu; Wu He; Shancang Li, "Internet of Things in Industries: A Survey," *Industrial Informatics, IEEE Transactions on* , Vol.10, No.4, pp.2233-2243
- [24] 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
- [25] Srinivasan, S.; Barchas, I.; Gorenberg, M.; Simoudis, E.2014.Venture Capital: Fueling the Innovation Economy. *Computer*, Vol.47, No.8, pp.40-47
- [26] 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
- [27] Chelliah, P.R.2014. Elucidating the Cloud Enterprise Architecture for Smarter Enterprises. *IT Professional* , Vol.16, No.6, pp.33,37
- [28] Zhou, Z-H, Chawla, N.V.; Yaochu J; Williams, G.J.2014. Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives [Discussion Forum]. *Computational Intelligence Magazine, IEEE* , Vol.9, No.4, pp.62-74