International Journal of Scientific Research & Engineering Trends
Volume 5, Issue 2, Mar-Apr-2019,ISSN (Online): 2395-566X

Survey on Big Data Concepts, Technology, Applications and Challenges

Asst. Prof. Sapna S. Kaushik
Department of Computer Science
NIE First Grade College
Mysuru , Karnataka, India
sapnakaushik123@gmail.com

Abstract **- Big data is a new driver of the world economic and societal changes. Incredible amounts of data is being generated by various organizations like hospitals, banks, e-commerce, retail and supply chain, social media and smart phones, Internet of Things and cloud computing etc. by virtue of digital technology . While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Big Data Analytics technologies. The voluminous data generated from the various sources can be processed and analyzed to support decision making. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Big Data Analytics is the process of examining large amounts of data (big data) to discover hidden patterns or unknown correlations. A key part of big data analytics is the need to collect, maintain and analyse enormous amounts of data efficiently and effectively. Analyzing Big Data is a challenging task as it contains huge dispersed, diverse and complex file systems which should be fault tolerant, flexible and scalable. There is an immense need of constructions, platforms, tools, techniques and algorithms to handle Big Data. The technologies used by big data application to handle the massive data are Hadoop, Map Reduce, Apache Hive, HBase,HDFS - Distributed File System.**

Keywords**- Big data characteristics, Hadoop Framework, HDFS,Map Reduce, Big data Analytics, Issues.**

I. INTRODUCTION

In digital world, data are generated from various sources and the fast transition in digital technologies has led to growth of big data. In general, it refers to the collection of large and complex datasets which are difficult to process using traditional database management tools or data processing applications. These are available in structured, semi-structured, and unstructured format in peta bytes and beyond.

Formally, it is defined from 3Vs to 4Vs. 3Vs refers to volume, velocity, and variety. Volume refers to the huge amount of data that are being generated everyday whereas velocity is the rate of growth and how fast the data are gathered for analysis.

Variety provides information about the types of data such as structured, unstructured, semi structured etc. The fourth V refers to veracity that includes availability and accountability. Veracity focuses on the quality of the data. This characterizes big data quality as good, bad, or undefined due to data inconsistency, incompleteness, ambiguity, latency, deception, and approximations [1].

Fig. 1 Schematic Representation of 3 V's of Big Data

1. Volume- This essentially concerns the large quantities of data that is generated continuously. Initially storing such data was problematic because of high storage costs. However with decreasing storage costs, this problem has been kept somewhat at bay as of now. However this is only a temporary solution and better technology needs to be developed. Smartphones, E-Commerce and social networking websites are examples where massive amounts of data are being generated.

2. Velocity- In what now seems like the pre-historic times, data was processed in batches. However this technique is only feasible when the incoming data rate is slower than the batch processing rate and the delay is much of a hindrance. At present times, the speed at which such colossal amounts of data are being generated is unbelievably high. Take Facebook [2] for

© 2019 IJSRET

example – it generates 2.7 billion like actions/day and 300 million photos amongst others roughly amounting to 2.5 million pieces of content in each day while Google processes over 1.2 trillion searches per year worldwide.

3. Variety- Documents to databases to excel tables to pictures and videos and audios in hundreds of formats, data is now losing structure. Structure can no longer be imposed like before for the analysis of data. Data generated can be of any type - structured, semi-structured or unstructured. The conventional form of data is structured data. For example text. Unstructured data can be generated from social networking sites, sensors and satellites. From the perspective of the information and communication technology, big data is a robust impetus to the next generation of information technology industries [3], which are broadly built on the third platform, mainly referring to big data, cloud computing, internet of things, and social business.

Implementing Big Data is a mammoth task given the large volume, velocity and variety. Big Data is a term encompassing the use of techniques to capture, process, analyze and visualize potentially large datasets in a reasonable timeframe not accessible to standard IT technologies. By extension, the platform, tools and software used for this purpose are collectively called Big Data technologies. Generally, Data warehouses have been used to manage the large dataset. In this case extracting the precise knowledge from the available big data is a foremost issue. Most of the presented approaches in data mining are not usually able to handle the large datasets successfully.

The key problem in the analysis of big data is the lack of coordination between database systems as well as with analysis tools . These challenges generally arise when we wish to perform knowledge discovery and representation for its practical applications. A fundamental problem is how to quantitatively describe the essential characteristics of big data [3]. Additionally, the study on complexity theory of big data will help understand essential characteristics and formation of complex patterns in big data, simplify its representation, gets better knowledge abstraction, and guide the design of computing models and algorithms on big data [3]. However, it is to be noted that all data available in the form of big data are not useful for analysis or decision making process. Industry and academia are interested in disseminating the findings of big data. This paper focuses on Big Data technologies, applications and challenges. So, to elaborate this, the paper is divided into following sections.

Sections 1 deals with the different technologies used in big data. Section 2 deals with the different areas where big data analytics is used. Section 3 provides an insight with challenges that arise during fine tuning of big data. Conclusion remarks are provided in section 4 to summarize outcomes.

II. BIG DATA TECHNOLOGIES

Big data environment calls for Magnetic, Agile, Deep (MAD) analysis skills, which differ from the aspects of a traditional Enterprise Data Warehouse (EDW) environment. First of all, traditional EDW approaches discourage the incorporation of new data sources until they are cleansed and integrated. Due to the ubiquity of data nowadays, big data environments need to be magnetic, thus attracting all the data sources, regardless of the data quality [4].

Big data storage requires an agile database, whose logical and physical contents can adapt in sync with rapid data evolution [5]. Current data analyses use complex statistical methods, and analysts need to be able to study enormous datasets by drilling up and down, a big data repository also needs to be deep, and serve as a sophisticated algorithmic runtime engine [4].

1. Hadoop : Framework

Apache Hadoop is an open-source software framework. It is used for distributed storage and distributed processing of very large data sets on computer clusters. It uses commodity hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common and should be automatically handled by the framework [6]. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Hadoop is a framework for performing big data analytics which provides reliability, scalability, and manageability by providing an implementation for the MapReduce paradigm. The base Apache Hadoop framework is composed of the following modules shown in Fig.1.

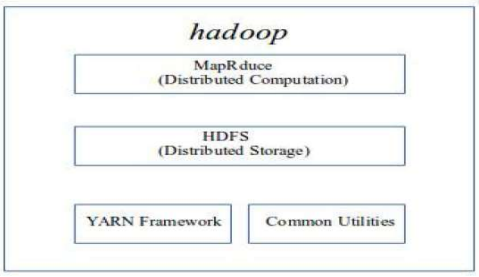

Fig. 2 Hadoop Framework Components.

2. Hadoop Common- It contains Java libraries and utilities needed. These libraries provide file system and OS level abstractions. It contains the necessary Java Archive files and scripts to start Hadoop.

3. **Hadoop Distributed File System (HDFS)**: It is a distributed file-system that stores data on commodity machines, providing very high bandwidth across the cluster. It provides high fault tolerance and native support of large data sets. It stores data on commodity hardware.
4. **Hadoop YARN**- This is a resource-management platform responsible for managing computing resources in clusters and using them for scheduling of users applications [7, 8].
5. **Hadoop MapReduce-** It is an implementation of the MapReduce programming model for large scale data processing. It is used for parallel processing of large datasets.

Hadoop consists of two main components: the HDFS for the big data storage, and MapReduce for big data analytics [9]. HDFS is highly fault tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets. The HDFS storage function provides a redundant and reliable distributed file system, which is optimized for large files, where a single file is split into blocks and distributed across cluster nodes. Additionally, the data is protected among the nodes by a replication mechanism, which ensures availability and reliability despite any node failures [10].

HDFS stores file system metadata and application data separately. HDFS stores metadata on a Dedicated server, called the Name Node. Application data are stored on other servers called Data Nodes. All servers are fully connected and communicate with each other using TCP-based protocols. There are two types of HDFS nodes: the Data Nodes and the Name Nodes. Data is stored in replicated file blocks across the multiple Data Nodes, and the Name Node acts as a regulator between the client and the Data Node, directing the client to the particular Data Node which contains the requested data [10].

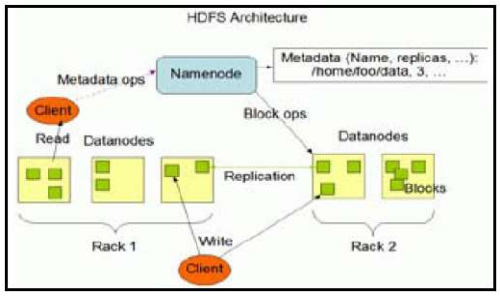

Fig. 3 HDFS Architecture.

The base node types for a Hadoop cluster are:
6. Name N.ode –The NameNode is the central location for information about the file system deployed in a Hadoop environment. Files and directories are represented on the Name Node by inodes, which record attributes like permissions, modification and access times, namespace and disk space quotas. The file content is split into large blocks (typically 128 megabytes) and each block of the file is independently replicated at multiple DataNodes (typically three). An environment can have one or two NameNodes, configured to provide minimal redundancy between the NameNodes. The NameNode[11] is contacted by clients of the Hadoop Distributed File System (HDFS) to locate information within the file system and provide updates for data they have added, moved, manipulated, or deleted. HDFS keeps the entire namespace in RAM.
7. Data Node –Data Nodes make up the majority of the servers contained in a Hadoop environment. Common Hadoop environments will have more than one Data Node, and oftentimes they will number in the hundreds based on capacity and performance needs. Each block replica on a Data Node is represented by two files in the local host's native file system.

The first file contains the data itself and the second file is block's meta data including checksums for the block data and the block's generation stamp. The size of the data file equals the actual length of the block and does not require extra space to round it up to the nominal block size as in traditional file systems. Thus, if a block is half full it needs only half of the space of the full block on the local drive. The Data Node [11] serves two functions: It contains a portion of the data in the HDFS and it acts as a computer platform for running jobs, some of which will utilize the local data within the HDFS.

IV. COMPONENTS OF HADOOP

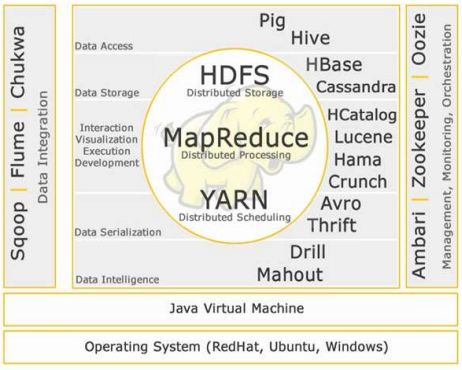

Fig. 4 Hadoop Ecosystem.

1. Pig- Pig is high-level platform where the MapReduce framework is created which is used with Hadoop platform. It is a high level data processing system where the data records are analyzed that occurs in high level language.

© 2019 IJSRET

2. Hive- It is application developed for data warehouse that provides the SQL interface as well as relational model. Hive infrastructure is built on the top layer of Hadoop that help in providing conclusion, and analysis for respective queries.
3. H Base - It is open source, distributed and Non relational database system implemented in Java. It runs above the layer of HDFS. It can serve the input and output for the Map Reduce in well-mannered structure.
4. H.Catalog - It is a table and storage management layer for Hadoop that enables users with different data processing tools.
5. Avro- It is a system that provides functionality of data serialization and service of data exchange. It is basically used in Apache Hadoop. These services can be used together as well as independently according the data records.
6. Mahout - It is a powerful, scalable machine-learning library that runs on top of Hadoop MapReduce.
7. Zookeeper- It is a centralization based service that provides distributed synchronization and provides group services along with maintenance of the configuration information and records.
8. Oozie- Oozie is a web-application that runs in a java servlet. Oozie use the database to gather the information of workflow which is a collection of actions. It manages the Hadoop jobs in a mannered way.
9. Sqoop- Sqoop is a command-line interface application that provides platform which is used for converting data from relational databases to Hadoop or vice versa.
10. Chukwa- Chukwa is a framework that is used for data collection and analysis to process and analyze the massive amount of logs. It is built on the upper layer of the HDFS and Map Reduce framework.

V. BIG DATA ANALYTIC PROCESSING

After the big data storage, comes the analytic processing. According to [12], there are four critical requirements for big data processing. The first requirement is fast data
loading. The second requirement is fast query processing. In order to satisfy the requirements of heavy workloads and real-time requests the data placement structure must be capable of retaining high query processing speeds as the amounts of queries rapidly increase.

The third requirement for big data processing is the highly efficient utilization of storage space. Finally, the fourth requirement is the strong adaptively to highly dynamic workload patterns. As big data sets are analyzed by different applications and users, for different purposes, and in various ways, the underlying system should be highly adaptive to unexpected dynamics in data processing, and not specific to certain workload patterns [12].

1. Map Reduce

Map Reduce is a parallel programming model, inspired by the "Map" and "Reduce" of functional languages, which is suitable for big data processing. It is the core of Hadoop, and performs the data processing and analytics functions [13]. The fundamental idea of MapReduce is breaking a task down into stages and executing the stages in parallel in order to reduce the time needed to complete the task [13].

The first phase of the MapReduce job is to map input values to a set of key/value pairs as output. The "Map" function accordingly partitions large computational tasks into smaller tasks, and assigns them to the appropriate key/value pairs [13]. Thus, unstructured data, such as text, can be mapped to a structured key/value pair, where, for example, the key could be the word in the text and the value is the number of occurrences of the word.
This output is then the input to the "Reduce" function. Reduce then performs the collection and combination of this output, by combining all values which share the same key value, to provide the final result of the computational task [13].

2. Map Reduce Architecture

Google created Map Reduce to process large amounts of unstructured or semi-structured data, such as web documents and logs of web page requests, on large clusters of commodity nodes. It produced various kinds of data such as inverted indices or URL access frequencies. The Map Reduce has three major parts, including Master, Map function and Reduce function. The Master is responsible for managing the back-end Map and Reduce functions and offering data and procedures to them [15, 16].

A Map Reduce application contains a workflow of jobs where each job makes two user-specified functions: Map and Reduce. The Map function is applied to each input record and produces a list of intermediate records. The Reduce function (also called Reducer) is applied to each group of intermediate records with the same key and produces a list of output records [14]. Map Reduce program is expected to be done on several computers and nodes when it is performed on Hadoop [17].

Therefore, a master node runs all the necessary services to organize the communication between Mappers and Reducers. An input file (or files) is separated into the same parts called input splits. They pass to the Mappers in which they work parallel together to provide the data contained within each split then each Reducer gathers the data partition by each Mapper, merges them, processes them, and produces the output file. An example of this data flow is shown in Fig. 1.

© 2019 IJSRET

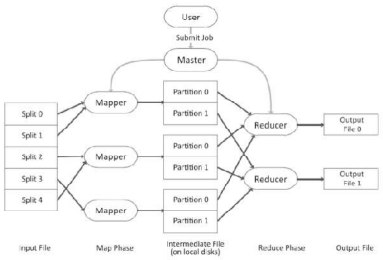

Fig. 4 Data flows in MapReduce Architecture

The main phases of MapReduce architecture are Mapper, Reducer, and shuffle which are presented below:

1. Mapper- a Mapper processes input data which are assigned by the master to perform some computation on this input and produce intermediate results in the form of key/value pairs [18].

2. Reducer- The Reduce function receives an intermediate key and a set of values of the key. It combines these values together to form a lesser set of values [19]. Information processing tasks benefit from parallel and distributed architecture with simply the programming of Map and Reduce methods.

- Map Reduce architecture has the ability to process Terabytes of data on PC clusters with handling failures.
- Most of the data recovery and excavating information can be taken into Map Reduce architecture, similar to the pattern based annotation algorithms. Distributed Grep is one of the basic examples for Map Reduce using Ontea pattern approach with regular expressions as well.

VI. BIG DATA ANALYTICS

The significance of big data lies in its ability to provide information and knowledge of value, upon which to base decisions. Big data is becoming an increasingly important asset for decision makers. Large volumes of highly detailed data from various sources such as scanners, mobile phones, loyalty cards, the web, and social media platforms provide the opportunity to deliver significant benefits to organizations. This is possible only if the data is properly analyzed to reveal valuable insights, allowing for decision makers to capitalize upon the resulting opportunities from the wealth of historic and real-time data generated through supply chains, production processes, customer behaviors, etc. [20]. This section focuses on some of the different areas where big data analytics is used, and how it can aid organizations across different sectors to gain valuable insights and enhance decision making.

1. Social Media Analytics

Social media has recently become important for social networking and content sharing. Yet, the content that is generated from social media websites is enormous and remains largely unexploited. Social media analytics is based on developing and evaluating informatics frameworks and tools in order to collect, monitor, summarize, analyze, as well as visualize social media data such as websites, blogs etc. and its uses in business purpose or decision making. Now a days Social Media is the best platform for understanding the real-time customer choice or intentions and sentiments, using social media business advertising, product marketing etc.

Furthermore, social media analytics facilitates understanding the reactions and conversations between people in online communities, as well as extracting useful patterns and intelligence from their interactions, in addition to what they share on social media websites. Applications of social media analytics are Behavior Analytics , Location-based interaction Analytics , Recommender systems development , Link prediction , Customer interaction and Analytics & marketing , Media use , Security , Social studies

Social Network Analysis (SNA) focuses on the relationships among social entities, as well as the patterns and implications of such relationships[21]. An SNA maps and measures both formal and informal relationships in order to comprehend what facilitates the flow of knowledge between interacting parties, such as who knows who, and who shares what knowledge or information with who and using what [22].

2. Content Base Analytics

Content Base Analytics means whatever data that store in social media back-end site. For example Facebook users store their data, photos, and videos on Facebook storage. Content-based predictive analytics recommender systems mostly match features (tagged keywords) among similar items and the user's profile to make recommendations.

3. Text Analytics

Text mining is used to analyze a document or set of documents in order to understand the content within and the meaning of the information contained. Text mining has become very important nowadays since most of the information stored, not including audio, video, and images, consists of text. While data mining deals with structured data, text presents special characteristics which basically follow a non-relational form [23]. Text analytics widely use in government, research, and business needs.

4. Sentiment Analytics

Sentiment analysis, or opinion mining, is becoming more and more important as online opinion data, such as blogs, product reviews, forums, and social data from social media sites like Twitter and Facebook, grow tremendously. Sentiment analysis focuses on analyzing and understanding emotions from subjective text patterns, and is enabled through text mining. It

© 2019 IJSRET

identifies opinions and attitudes of individuals towards certain topics, and is useful in classsifying viewpoints as positive or negative. Sentiment analysis uses natural language processing and text analytics in order to identify and extract information by finding words that are indicative of a sentiment, as well as relationships between words, so that sentiments can be accurately identified [24].

5. Audio Analytics Audio analytics is the process of compressing data and packaging the data in to single format called audio. Audio Analytics refers to the extraction of meaning and information from audio signals for Analysis. There are two way to represent the audio Analytics is 1) Sound Representation 2) Raw Sound Files. Audio analytics is used in Surveillance application , Detection of Threats, Tele-monitoring System

6. Video Analytics

Video is a major issue when considering big data. Videos and images contribute to 80 % of unstructured data. Now a days, CCTV cameras are the one form of digital information and surveillance. All these information is stored and processed for further use. Video analytics is used in accident cases, schools, traffic police, business, security etc, investigation , Business Intelligence, Target and Scene Analytics, Direction Analytics. Advanced Data Visualization (ADV) and visual discovery has emerged as a powerful technique to discover knowledge from data. ADV combines data analysis methods with interactive visualization to enable comprehensive data exploration. It is a data driven exploratory approach that fits well in situations where analysts have little knowledge about the data [25].

With the generation of more and more data of high volume and complexity, an increasing demand has arisen for ADV solutions from many application domains [26]. ADV can enable faster analysis, better decision making, and more effective presentation and comprehension of results by providing interactive statistical graphics and a point-and-click interface . Furthermore, ADV is a natural fit for big data since it can scale its visualizations to represent thousands or millions of data points, unlike standard pie, bar, and line charts.

Moreover, it can handle diverse data types, as well as present analytic data structures that aren't easily flattened onto a computer screen, such as hierarchies and neural nets. Additionally, most ADV tools and functions can support interfaces to all the leading data sources, thus enabling business analysts to explore data widely across a variety of sources in search of the right analytics dataset, usually in real-time [27].

VII. CHALLENGES OF BIG DATA

One of the key set of challenges [28] faced in today's tight market is the need to find and analyze the required data at the least speed possible. However with exponentially growing amount of data, speed becomes a major issue as analyzing such sheer volumes of data in detail to find out required output becomes more and more tedious. It is not only the quantity of data but also discovering the data according to the appropriateness of the project which is a Herculean task.

Elimination of out-of-context data is an essential objective. Even if in-context data retrieved at a high speed is achieved, the quality of data may be compromised if it is not accurate or timely. As a result of this, appropriate results of the project may not be published. Another zone of challenges involves those relating to the vulnerability and security of Big Data. Breaches of privacy, especially with data relating to individuals and organizations have been a topic of serious concern. One solution has been to anonymize data by removing identifiers which could be used to pinpoint particular individuals thus compromising their privacy Organizations dealing with big data need to take this issue in their stride and make sure that the data storage and location be made heavily protected so that it is not misused.

VIII. CONCLUSION

In this paper the innovative topic of big data, which has recently gained lots of interest due to its perceived unprecedented opportunities and benefits has been examined. In recent years data are generated at a dramatic pace. Analyzing these data is challenging for a general man. In this paper we survey the various characteristics of big data, architecture, tool used to analyze big data,various areas where big data analytics is used and its challenges. Big Data is an evolving field, where much of the research is yet to be done.

REFERENCES

1. TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: TechAmerica Reports, pp. 1–40 (2012)
2. R. Kitchin, Big Data, new epistemologies and paradigm shifts, Big Data Society, 1(1) (2014), pp.1-12.
3. X. Jin, B. W.Wah, X. Cheng and Y. Wang, Significance and challenges of big data research, Big Data Research, 2(2) (2015), pp.59-64.
4. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492 (2009)

© 2019 IJSRET

5. Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F.B., Babu, S.: Starfish: A Self-tuning System for Big Data Analytics. In: Proceedings of the Conference on Innovative Data Systems Research, pp. 261–272 (2011)
6. "Welcome to Apache Hadoop!". hadoop.apache.org. Retrieved 2016-08-25.
7. Resource (Apache Hadoop Main 2.5.1 API)". apache.org. Apache Software Foundation. 2014-09-12. Retrieved 2014-09-30.
8. Murthy, Arun (2012-08-15). "Apache Hadoop YARN Concepts and Applications". hortonworks.com. Hortonworks. Retrieved 2014-09-30.
9. EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)
10. Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)
11. Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler: "The Hadoop Distributed File System" IEEE,2010.
12. He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., Xu, Z.: RCFile: A Fast and Spaceefficient Data Placement Structure in MapReduce-based Warehouse Systems. In: IEEE International Conference on Data Engineering (ICDE), pp. 1199–1208 (2011)
13. Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011).
14. Lee, D., Kim, J.-S., Maeng, S.: Large-scale incremental processing with MapReduce. Futur. Gener. Comput. Syst. 36, 66–79 (2014)
15. Wu, T.-Y., Chen, C.-Y., Kuo, L.-S., Lee,W.-T., Chao, H.-C.: Cloud-based image processing system with priority-based data distribution mechanism. Comput. Commun. **35**, 1809 – 1818 (2012).
16. Senger, H., Gil-Costa, V., Arantes, L., Marcondes, C.A.C., Marín, M., Sato, L.M., et al.: BSP cost and scalability analysis for MapReduce operations. Concurr. Comput. Pract. Exp. **28**, 2503–2527 (2016)
17. Idris, M., Hussain, S., Ali, M., Abdulali, A., Siddiqi, M.H., Kang, B.H., et al.: Context-aware scheduling in MapReduce: a compact review. Concurr. Comput. Pract. Exp. **27**, 5332–5349 (2015).
18. Lee, C.-W., Hsieh, K.-Y., Hsieh, S.-Y., Hsiao, H.-C.: A dynamic data placement strategy for Hadoop in heterogeneous environments. Big Data Res. **1**, 14–22 (2014).
19. Aridhi, S., d'Orazio, L., Maddouri, M., Mephu Nguifo, E.: Density-based data partitioning strategy to approximate large-scale subgraph mining. Inf. Syst. **48**, 213–223 (2015).
20. Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012).
21. Van der Valk, T., Gijsbers, G.: The Use of Social Network Analysis in Innovation Studies: Mapping Actors and Technologies. Innovation: Management, Policy & Practice 12(1), 5–17 (2010).
22. Serrat, O.: Social Network Analysis. Knowledge Network Solutions 28, 1–4 (2009).
23. Sanchez, D., Martin-Bautista, M.J., Blanco, I., Torre, C.: Text Knowledge Mining: An Alternative to Text Data Mining. In: IEEE International Conference on Data Mining Workshops, pp. 664–672 (2008).
24. Mouthami, K., Devi, K.N., Bhaskaran, V.M.: Sentiment Analysis and Classification Based on Textual Reviews. In: International Conference on Information Communication and Embedded Systems (ICICES), pp. 271–276 (2013).
25. Shen, Z., Wei, J., Sundaresan, N., Ma, K.L.: Visual Analysis of Massive Web Session Data. In: Large Data Analysis and Visualization (LDAV), pp. 65–72 (2012).
26. Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H., Keim, D.: Visual Analytics for the Big Data Era A Comparative Review of State-of-the-Art Commercial Systems. In: IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 173–182 (2012)
27. Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)
28. SAS, The power to know, —Five Big Data Challenges And how to overcome them with visual analytics.

© 2019 IJSRET