Journal of Advanced Research in Technology and Management Sciences
Volume: 01, Issue: 01, January 2019
Available online at: http://www.jartms.org

BIG DATA ANALYSIS FOR BETTER PERFORMANCE IN RECENT PLATFORMS AND ITS CHALLENGES

Dr.M.Dhasaratham[1], Dr.K.Sarangam[2]
[1,2]Assistant Professor, [1] Bule Hora University, Ethiopia.
[2] Department of CS, Brilliant Institute of Engineering and Technology, Hyderabad
E-mail: dasarath.m@gmail.com, E-mail: k.sarangam@gmail.com

Abstract: In gift days call manufacturers have monumental knowledge in their hands for knowledge analysis. Knowledge sets that are huge and even have high rate of variations which might be troublesome to handle with standard tools by victimization any technique that's an enormous knowledge. Victimization huge knowledge could be an information knowledge answer while not knowing that we tend to can't extract any knowledge sets so; an intensive study is required during this technology. It's going to incorporates multiple levels which can need effort for analysing and reading the info. huge knowledge analysis could be a current space of analysis and development. IoT could be a new platform wherever we tend to use this to use. In addition, range of analysis programs is current on this huge knowledge which can need additional solutions.

Keywords*:-Big knowledge analytics; large data; Structured data; Unstructured Data; IoT*

Introduction: In digital surroundings, knowledge is generated from numerous sources of technologies that have LED to growth of massive knowledge. It provides organic process breakthroughs in several fields with an assortment of huge datasets. The approach within which assortment of huge and complicated datasets become troublesome to method victimization ancient management tools or processing applications. These are accessible in a structured, semi-structured, and unstructured format in petabytes and on the far side. Formally, it's outlined from 3Vs to 4Vs. 3Vs refers to volume, velocity, and selection. Volume refers to the large quantity of knowledge that are being generated every day whereas rate is that the rate of growth and the way quick the info are gathered for being analysis. Selection provides info regarding the categories of knowledge like structured, unstructured, semi-structured etc. The fourth V refers to truthfulness that has accessibility and answerableness. The prime objective of massive knowledge analysis is to method knowledge of high volume, velocity, variety, and truthfulness victimization numerous ancient and machine intelligent techniques. The subsequent Figure one refers to the definition of massive knowledge. But the actual definition for large knowledge isn't outlined and there's believed that it's drawback specific. This can facilitate the North American nation in getting increased deciding, insight discovery and improvement whereas being innovative and efficient. It's expected that the expansion of massive knowledge is calculable to succeed in twenty-five billion by 2015. From the angle of the knowledge and communication technology, huge knowledge could be a strong motivation to the consequent generation of knowledge technology industries, that are generally engineered on the third platform, principally touching on huge knowledge, cloud computing, web of things, and social business. The key drawback within the analysis of massive knowledge is that the lack of coordination between information systems additionally likes analysis tools like data processing and applied math analysis. These challenges usually arise once we would like to perform information discovery and illustration for its sensible applications. An elementary drawback is a way to quantitatively describe the essential characteristics of massive knowledge. in addition, the study on quality theory of huge knowledge can facilitate perceive essential characteristics and formation of advanced patterns in big knowledge, modify its illustration, gets higher information abstraction, and guide the planning of computing models and algorithms on huge knowledge. Abundant analysis was disbursed by numerous researchers on huge knowledge and its trends. The fundamental objective of this paper is to explore the potential impact of massive knowledge challenges, its analysis problems, and numerous tools related to it. As a result, this text provides a platform to explore huge knowledge at totally different stages. in addition, we tend to state open analysis problems in huge knowledge. This paper is split into the following sections.

Challenges: In Recent years big data has been accumulated in several domains like health care, retail, bio-chemistry, and other interdisciplinary scientific researches. Web-based applications encounter big data frequently, such as social computing, internet text and documents, and inter-net search indexing. Social computing includes social net-work analysis, online communities, recommender systems, reputation systems, and prediction markets where as internet search indexing includes ISI, IEEE Xplorer, Scopus, Thomson, Reuters, etc.

Considering this advantages of big data it provides a new opportunities in the knowledge processing tasks for the upcoming researchers. However opportunities always follow some challenges. To handle the challenges we need to know various computational complexities, information security, and computational method, to analyze big data. For example, many statistical methods that perform well for small data size do not scale to huge data. Here the challenges of big data analytics are classified into four broad categories namely data storage and analysis; knowledge discovery and computational complexities; scalability and visualization of data; and information security. We discuss these issues briefly in the following subsections.

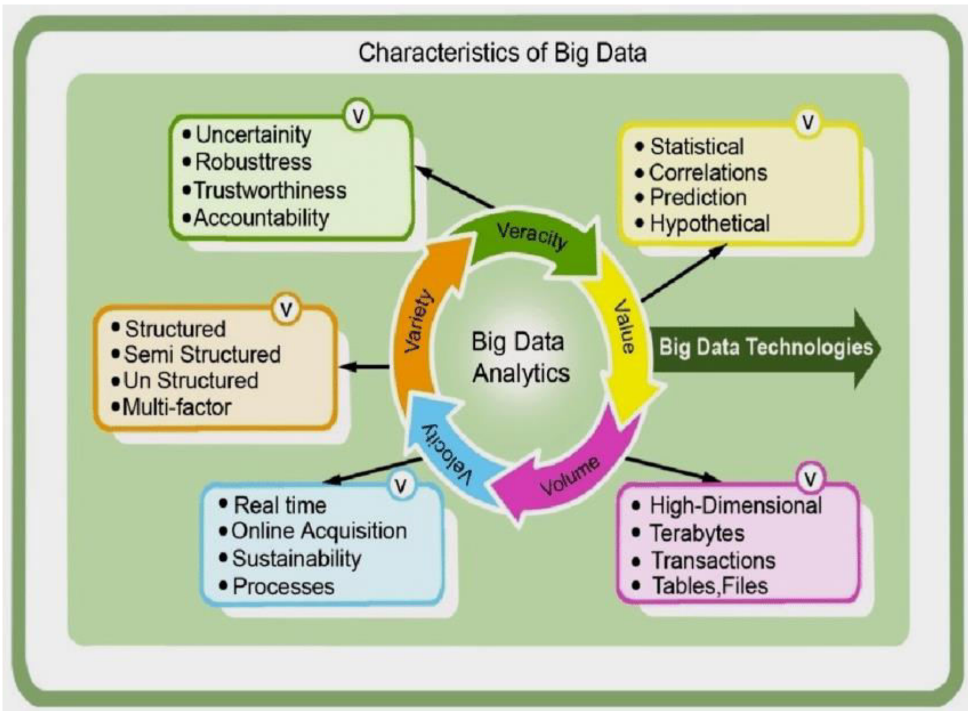

Figure 1: Characteristics of Big Data.

A. Storage Analysis: In recent years the size of data has grown exponentially by various means such as mobile devices, sensor technologies, remote sensing, radio frequency identification readers etc. These data are stored on spending much cost whereas they ignored or deleted finally because there is no enough space to store them. Therefore, the first challenge for big data analysis is storage mediums and higher input/output speed. In such cases, the data accessibility must be on the top priority for the knowledge discovery and representation. The prime reason is being that, it must be accessed easily and promptly for further analysis. In past decades, analyst use hard disk drives to store data but, it slower random input/output performance than sequential input/output. To overcome this limitation, the concept of solid state drive (SSD) and phrase change memory (PCM) was introduced. However the available storage technologies cannot possess the required performance for processing big data. Another challenge with Big Data analysis is attributed to diversity of data with the ever growing of datasets, data mining tasks has significantly increased. Additionally data reduction, data selection, feature selection is an essential task especially when dealing with large datasets. This presents an unprecedented challenge for researchers. It is because, existing algorithms may not always respond in an adequate time when dealing with these high dimensional data. Automation of this process and developing new machine learning algorithms to ensure consistency is a major challenge in recent years. The major challenge in this case is to pay more attention for designing storage systems and to elevate efficient data analysis tool that provide guarantees on the output when the data comes from different sources. Furthermore, design of machine learning algorithms to analyze data is essential for improving efficiency and scalability.

B. **Arithmetic Complexities:** Knowledge discovery and representation is a prime issue in big data. It includes a number of sub fields such as authentication, archiving, management, preservation, information retrieval, and representation. Many hybridized techniques are also developed to process real life problems. All these techniques are problem dependent. Further some of these techniques may not be suitable for large datasets in a sequential computer. At the same time some of the techniques has good characteristics of scalability over parallel computer. Since the size of big data keeps increasing exponentially, the available tools may not be efficient to process these data for obtaining meaningful information. The most popular approach in case of large dataset management is data warehouses and data marts. Data warehouse is mainly responsible to store data that are sourced from operational systems whereas data mart is based on a data warehouse and facilitates analysis.

Analysis of large dataset requires more computational complexities. The major issue is to handle inconsistencies and uncertainty present in the datasets. In general, systematic modeling of the computational complexity is used. It may be difficult to establish a comprehensive mathematical system that is broadly applicable to Big Data. But a domain specific data analytics can be done easily by understanding the particular complexities. A series of such development could simulate big data analytics for different areas. Much research and survey has been carried out in this direction using machine learning techniques with the least memory requirements. The basic objective in this research is to minimize computational cost processing and complexities. However, current big data analysis tools have poor performance in handling computational complexities, uncertainty, and inconsistencies. It leads to a great challenge to develop techniques and technologies that can deal computational complexity, uncertainty and inconsistencies in an effective manner.

C. **Data Scalability And Visualization**: The most important challenge for big data analysis techniques is its scalability and security. In the last decades researchers have paid attentions to accelerate data analysis and its speed up processors followed by Moore’s Law. For the former, it is necessary to develop sampling, on-line, and multi resolution analysis techniques. Incremental techniques have good scalability property in the aspect of big data analysis. As the data size is scaling much faster than CPU speeds, there is a natural dramatic shift in processor technology being embedded with increasing number of cores. This shift in processors leads to the development of parallel computing. Real time applications like navigation, social networks, finance, internet search, timeliness etc. requires parallel computing. We can observe that big data have produced many challenges for the developments of the hardware and software which leads to parallel computing, cloud computing, distributed computing, visualization process, scalability. To over-come this issue, we need to correlate more mathematical models to computer science.

D. Data Safety: In big data analysis massive amount of data are correlated, analyzed, and mined for meaningful patterns. All organizations have different policies to safe guard their sensitive information. Preserving sensitive information is a major issue in big data analysis. There is a huge security risk associated with big data. Therefore, information security is becoming a big data analytics problem. Security of big data can be enhanced by using the techniques of authentication, authorization, and encryption. Therefore, attention has to be given to develop a multilevel security policy model and prevention system. Although much research has been carried out to secure big data but it requires lot of improvement. The major challenge is to develop a multi-level security, privacy preserved data model for big data.

III. Recent Research In Big Data And Its Problems: Big data analytics and data science are becoming the research focal point in industries and academia. Data science aims at researching big data and knowledge extraction from data. Applications of big data and data science include information science, uncertainty modelling, uncertain data analysis, machine learning, statistical learning, pattern recognition, data warehousing, and signal processing. Effective integration of technologies and analysis will result in predicting the future drift of events. Main focus of this section is to discuss open research issues in big data analytics. The research issues pertaining to big data analysis are classified into three broad categories namely internet of things (IoT), cloud computing, bio inspired computing, and quantum computing. However it is not limited to these issues.

IoT for Big Data Analytics: Internet has restructured global interrelations, the art of businesses, cultural revolutions and an unbelievable number of personal characteristics. Currently, machines are getting in on the act to control innumerable autonomous gadgets via internet and create Internet of Things (IoT). Thus, appliances are becoming the user of the internet, just like humans with the web browsers. Internet of Things is attracting the attention of recent researchers for its most promising opportunities and challenges. It has an imperative economic and societal impact for the future construction of information, network and communication technology. The new regulation of future will be eventually; everything will be connected and intelligently controlled. The concept of IoT is becoming more pertinent to the realistic world due to the development of mobile de-vices, embedded and ubiquitous communication technologies, cloud computing, and data analytics. Moreover, IoT presents challenges in combinations of volume, velocity and variety. In a broader sense, just like the internet, Internet of Things enables the devices to exist in a myriad of places and facilitates applications ranging from trivial to the crucial. Conversely, it is still mystifying to understand IoT well, including definitions, content and differences from other similar concepts. Several diversified technologies such as computational intelligence, and big data can be incorporated together to improve the data management and knowledge discovery of large scale automation applications.

Knowledge acquisition from IoT data is the biggest challenge that big data professional are facing. Therefore, it is essential to develop infrastructure to analyze the IoT data. An IoT device generates continuous streams of data and the re-searchers can develop tools to extract meaningful information from these data using machine learning techniques. Under-standing these streams of data generated from IoT devices and analyzing them to get meaningful information is a challenging issue and it leads to big data analytics. Machine learning algorithms and computational intelligence techniques is the only solution to handle big data from IoT prospective. Key technologies that are associated with IoT are also discussed in many research papers. Figure 2 depicts an overview of IoT big data and knowledge discovery process. Knowledge exploration system has originated from theories of human information processing such as frames, rules, tagging, and semantic networks.

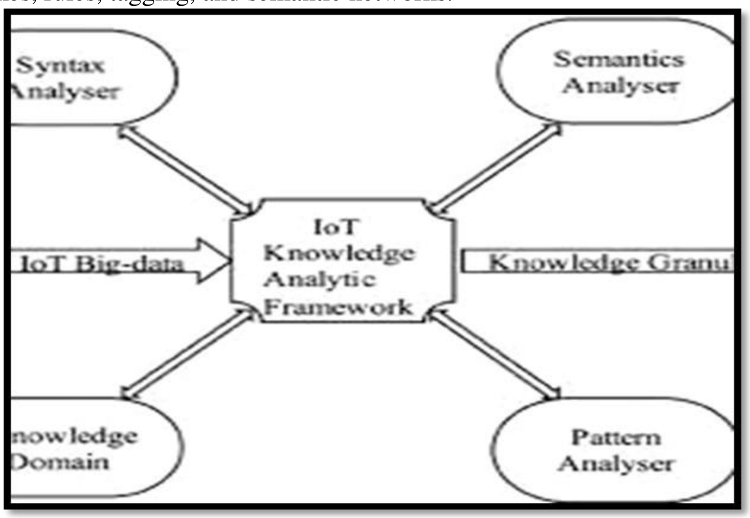

Figure 2: IoT Big Data Knowledge Discovery

In general, it consists of four segments such as knowledge acquisition, knowledge base, knowledge dissemination, and knowledge application. In knowledge acquisition phase, knowledge is discovered by using various traditional and computational intelligence techniques, the discovered knowledge is stored in knowledge bases and expert systems are generally designed based on the discovered knowledge. Knowledge dissemination is important for obtaining meaningful information from the knowledge bases and expert systems are generally designed based on the discovered knowledge. Knowledge dissemination is important for obtaining meaningful information from the knowledge base. Knowledge extraction is a process that searches documents, knowledge within documents as well as knowledge bases. The final phase is to apply discovered knowledge in various applications. It is the ultimate goal of knowledge discovery. The knowledge exploration system is necessarily iterative with the judgement of knowledge application. There are many issues, discussions, and researches in this area of knowledge exploration. It is beyond the scope of this survey paper. For better visualization, knowledge exploration system is depicted in Figure 3.

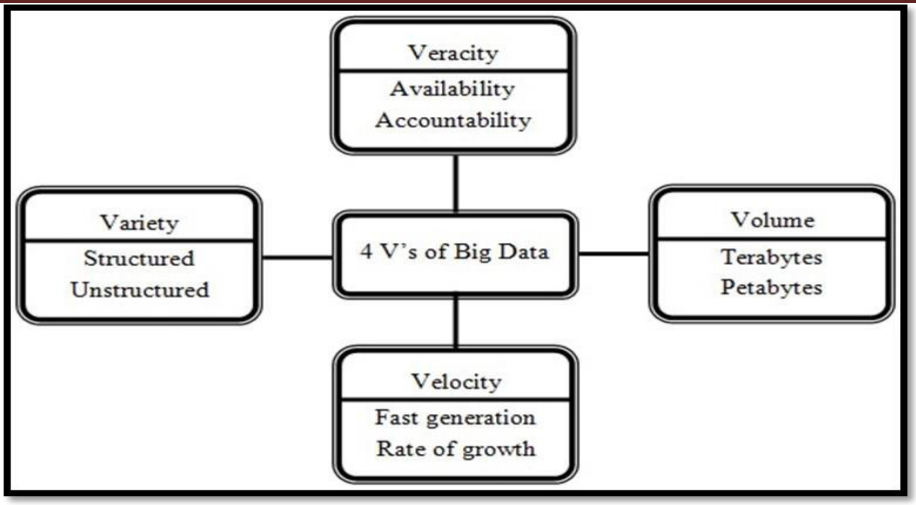

Figure 3: IoT Knowledge Exploration System

Cloud Computing for Big Data Analytics: The development of virtualization technologies has made Supercomputing more accessible and affordable. Computing infrastructures that are hidden in virtualization software make systems to behave like a true computer, but with the flexibility of specification details such as number of processors, disk space, memory, and operating system. The use of these virtual computers is known as cloud computing which has been one of the most robust big data technique. Big Data and cloud computing technologies are developed with the importance of developing a scalable and on demand availability of resources and data. Cloud computing harmonize massive data by on demand access to configurable computing resources through virtualization techniques. The benefits of utilizing the Cloud computing include offering resources when there is a demand and pay only for the resources which is needed to develop the product. Simultaneously, it improves availability and cost reduction. Open challenges and research issues of big data and cloud computing are discussed in detail by many re-searchers which highlights the challenges in data management, data variety and velocity, data storage, data processing, and resource management [16]. So Cloud computing helps in developing a business model for all varieties of applications with infrastructure and tools. Big data application using cloud computing should support data analytic and development. The cloud environment should provide tools that allow data scientists and business analysts to interactively and collaboratively explore knowledge acquisition data for further processing and extracting fruitful results. This can help to solve large applications that may arise in various domains.

In addition to this, cloud computing should also enable scaling of tools from virtual technologies into new technologies like spark, R, and other types of big data processing techniques. Big data forms a framework for discussing cloud computing options. Depending on special need, user can go to the Market place and buy infrastructure services from cloud service providers such as Google, Amazon, IBM, software as a service (SaaS) from a whole crew of companies such as Net Suite, Cloud9, Job science, etc. Another advantage of cloud computing is cloud storage which provides a possible way for storing big data. The obvious one is the time and cost that are needed to upload and download big data in the cloud environment. Else, it becomes difficult to control the distribution of computation and the underlying hardware. But, the major issues are privacy concerns relating to the hosting of data on public servers, and the storage of data from human studies. All these issues will take big data and cloud computing to a high level of development.

Computing On Bio-Data Analysis: Bio inspired computing is a technique inspired my nature to address complex real world problems. Biological systems are self- organized without a central control. A bio-inspired cost minimization mechanism search and find the optimal data service solution on considering cost of data management and service maintenance. These techniques are developed by biological molecules such as DNA and proteins to conduct computational calculations involving storing, retrieving, and processing of data. A significant feature of such computing is that it integrates biologically derived materials to perform computational functions and receive intelligent performance. These systems are more suitable for big data applications. Huge amount of data are generated from variety of resources across the web since the digitization. Analyzing these data and categorizing into text, image and video, etc. will require lot of intelligent analytics from data scientists and big data professionals. Proliferations of technologies are emerging like big data, IoT, cloud computing, bio inspired computing etc.

Whereas equilibrium of data can be done only by selecting right platform to analyze large and furnish cost effective results. Bio-inspired computing techniques serve as a key role in intelligent data analysis and its application to big data. These algorithms help in performing data mining for large datasets algorithms help in performing data mining for large datasets due to its optimization application. The most advantage is its simplicity and their rapid convergence to optimal solution while solving service provision problems. Some applications to this end using bio inspired computing were discussed in detail by Cheng et al.

Quantum Computing for Big Data Analysis: A quantum computer has memory that is exponentially larger than its physical size and can manipulate an exponential set of inputs simultaneously. This exponential improvement in computer systems might be possible. If a real quantum computer is available now, it could have solved problems that are exceptionally difficult on recent computers, of course today's big data problems. The main technical difficulty in building quantum computer could soon be possible. Quantum computing provides a way to merge the quantum mechanics to process the information. In traditional computer, information is presented by long strings of bits which encode either a zero or a one. On the other hand a quantum computer uses quantum bits or qubits. The difference between qubit and bit is that, a qubit is a quantum system that encodes the zero and the one into two distinguishable quantum states. Therefore, it can be capitalized on the phenomena of superposition and entanglement. For example, 100 qubits in quantum systems require 2100 complex values to be stored in a classic computer system. It means that many big data problems can be solved much faster by larger scale quantum computers compared with classical computers. Hence it is a challenge for this generation to build a quantum computer and facilitate quantum computing to solve big data problems.

PROCESSING TOOLS: Large numbers of tools are available to process big data. In this section, we discuss some current techniques for analysing big data with emphasis on three important engineering tools namely Map Reduce, Apache Spark, and Storm. Most of the available tools concentrate on batch processing, steam processing and interactive analysis. Most batch processing tools are based on the Apache Hadoop infrastructure such as Mahout and Dryad. Stream data applications are mostly used for real time analytic. Some examples of large scale streaming platform are Strom and Splunk. The interactive analysis process allows users to directly interact in real time for their own analysis. For example Dremel and Apache Drill are the big data plat-forms that support interactive analysis. These tools help us in developing the big data projects. A fabulous list of big data tools and techniques is also discussed by much researchers [6]. The typical work flow of big data project discussed by Huang et al is highlighted in this section and is depicted in Figure 4.

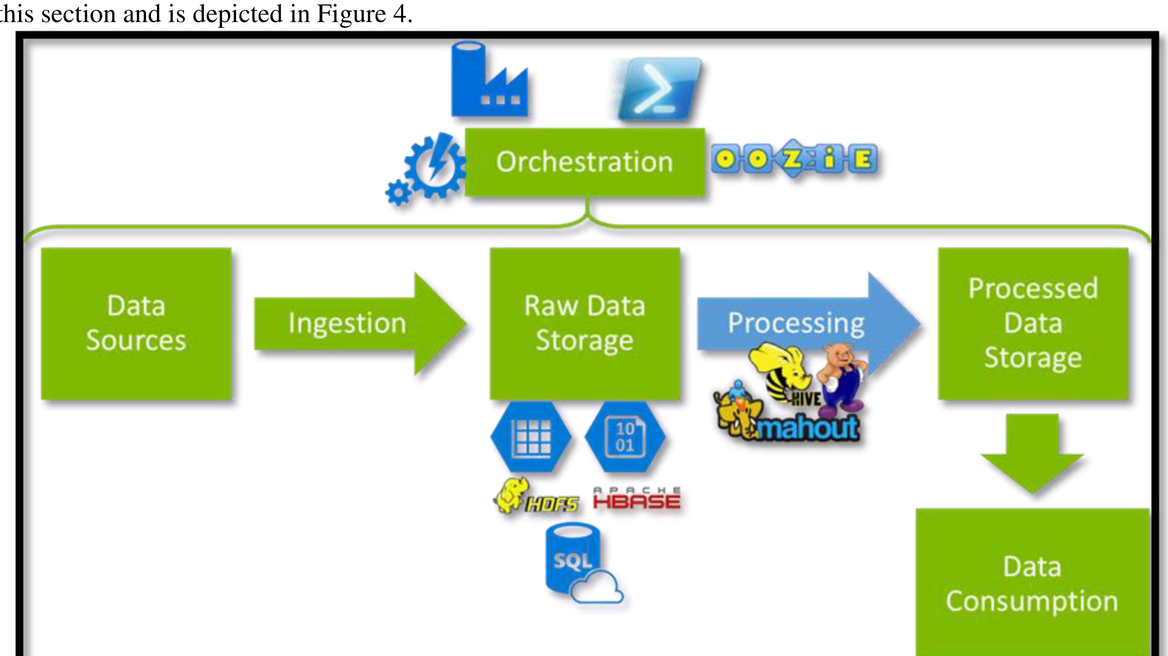

Figure 4: Work Flow of Big Project

Apache Mahout: Apache mahout aims to provide scalable and commercial machine learning techniques for large scale and intelligent data analysis applications. Core algorithms of mahout including clustering, classification, pattern mining, regression, dimensionality reduction, evolutionary algorithms, and batch based collaborative filtering run on top of Hadoop platform through map reduce framework. The goal of mahout is to build a vibrant, responsive, diverse community to facilitate discussions on the project and potential use cases. The basic objective of Apache Mahout is to provide a tool for elevating big challenges. The different companies those who have implemented scalable machine learning algorithms are Google, IBM, Amazon, Facebook, and Twitter.

Apache Spark: Apache spark is an open source big data processing frame- work built for speed processing, and sophisticated analytics. It is easy to use and was originally developed in 2009 in UC Berkeleys AMPLab. It was open sourced in 2010 as an Apache project. Spark lets you quickly write applications in java, scala, or python. In addition to map reduce operations, it supports SQL queries, streaming data, machine learning, and graph data processing. Spark runs on top of existing hadoop distributed file system (HDFS) infrastructure to provide enhanced and additional functionality. Spark consists of components namely driver program, cluster manager and worker nodes. The driver program serves as the starting point of execution of an application on the spark cluster. The cluster manager allocates the resources and the worker nodes to do the data processing in the form of tasks. Each application will have a set of processes called executors that are responsible for executing the tasks. The major advantage is that it provides support for deploying spark applications in an existing hadoop clusters. Figure 5 depicts the architecture diagram of Apache Spark. The various features of Apache Spark are listed below:

The prime focus of spark includes resilient distributed datasets (RDD), which store data in-memory and provide fault tolerance without replication. It supports iterative computation, improves speed and resource utilization. The foremost advantage is that in addition to Map Reduce it also supports streaming data, machine learning and graph algorithms. Another advantage is that user can run the application program in different languages such as Java, R, Python, or Scala. This is possible as it comes with higher-level libraries for advanced analytics. These standard libraries increases developer productivity and can be seamlessly combined to create complex work-flows Spark helps to run an application in Hadoop, Cluster, up to 100 times faster in memory, and 10 times faster when running on disk.

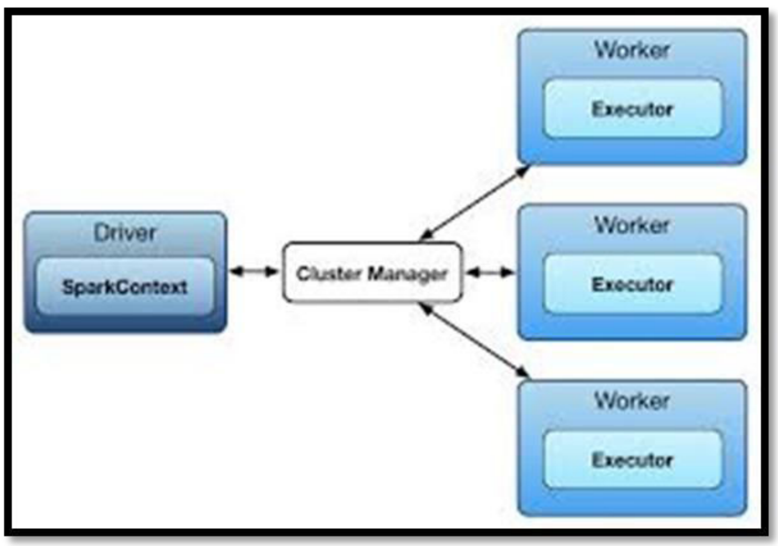

Figure 5 Architecture of Apache Spark

It is possible because of the reduction in number of read or writes operations to disk. It is written in Scala programming language and runs on java virtual machine (JVM) environment. Additionally, it supports java, python and R for developing applications using Spark.

Dryad: It is another popular programming model for implementing parallel and distributed programs for handling large context bases on dataflow graph. It consists of a cluster of computing nodes, and a user use the resources of a computer cluster to run their program in a distributed way. Indeed a dryad user uses thousands of machines, each of them with multiple processors or cores. The major advantage is that users do not need to know anything about concurrent programming. A dryad application runs a computational directed graph that is composed of computational vertices and communication channels. Therefore, dryad provides a large number of functionality including generating of job graph, scheduling of the machines for the available processes, transition failure handling in the cluster, collection of performance metrics, visualizing the job, invoking user defined policies and dynamically updating the job graph in response to these policy decisions without knowing the semantics of the vertices.

Storm: Storm is a distributed and fault tolerant real time computation system for processing large streaming data. It is specially designed for real time processing in contrasts with hadoop which is for batch processing. Additionally, it is also easy to set up and operate, scalable, fault-tolerant to provide competitive performances. The storm cluster is apparently similar to hadoop cluster. On storm cluster users run different topologies for different storm tasks whereas hadoop platform implements map reduce jobs for corresponding applications. There are number of differences between map reduce jobs and topologies. The basic differences is that map reduce job eventually finishes whereas a topology process messages all two kinds of nodes such as master node and worker node. The master node and worker node implement two kinds of roles such as nimbus and supervisor respectively. The two roles have similar functions in accordance with job tracker and task tracker of map reduce framework. Nimbus is in charge of distributing code across the storm cluster, scheduling and assigning tasks to worker nodes, and monitoring the whole system. The supervisor compiles tasks as assigned to them by nimbus. In addition, it start and terminate the process as necessary based on the instructions of nimbus the whole computational technology is partitioned and distributed to a number of worker process and each worker process implements a part of the topology.

Conclusion: Data analysis challenges are more now days for a common man. To this end in this paper, we survey the various research issues, challenges, and tools used to analyze these big data. From this survey, it is understood that every big data platform has its individual focus some of them are designed for batch processing whereas some are good at real-time analytic. Each big data platform also has specific functionality. Different techniques used for the analysis include statistical analysis, machine learning, data mining, intelligent analysis, cloud computing, quantum computing, and data stream processing. We believe that in future researchers will pay more attention to these techniques to solve problems of big data effectively and efficiently.

The amount of data collected from various applications all over the world across a wide variety of fields today is expected to double every two years. It has no utility unless these are analyzed to get useful information. This necessitates the development of techniques which can be used to facilitate big data analysis. The development of powerful computers is a boon to implement these techniques leading to automated systems. The transformation of data into knowledge is by no means an easy task for high performance large-scale data processing, including exploiting parallelism of current and upcoming computer architectures for data mining. Moreover, these data may involve uncertainty in many different forms. Many different models like fuzzy sets, rough sets, soft sets, neural networks, their generalizations and hybrid models obtained by combining two or more of these models have been found to be fruitful in representing data. These models are also very much fruitful for analysis. More often than not, big data are reduced to include only the important characteristics necessary from a particular study point of view or depending upon the application area. So, reduction techniques have been developed. Often the data collected have missing values. These values need to be generated or the tuples having these missing values are eliminated from the data set before analysis. More importantly, these new challenges may comprise, sometimes even deteriorate, the performance, efficiency and scalability of the dedicated data intensive computing systems. The later approach sometimes leads to loss of information and hence not preferred. This brings up many research issues in the industry and research community in forms of capturing and accessing data effectively. In addition, fast processing while achieving high performance and high throughput, and storing it efficiently for future use is another issue. Further, programming for big data analysis is an important challenging issue. Expressing data access requirements of applications and designing programming language abstractions to exploit parallelism are an immediate need. Additionally Machine learning concepts and tools are gaining popularity among researchers to facilitate meaningful results from these concepts. Research in the area of machine learning tools of big data are started recently needs drastic change to adopt it. We argue that while each other of the tool has their advantages and limitations, more efficient tools can be developed for dealing with problems inherent to big data. The efficient tools to be developed must have provision to handle noisy and imbalance data, uncertainty and inconsistency, and missing values.

REFERENCES

[1] M. K. Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, International Journal of Application or Innovation inEngineering &Management, 2(8) (2015), pp.228-232.
[2] C. Lynch, Big data: How do your data grow? Nature, 455 (2008), pp.28-29.
[3] X. Jin, B. W.Wah, X. Cheng and Y. Wang, Significance and challenges or big data research, Big Data Research, 2(2)(2015), pp.59-64.
[4] C. L. Philip, Q. Chen and C. Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, Information Sciences 275 (2014), pp.314-347.

[5] K. Kambatla, G. Kollias, V. Kumar and A. Gram, Trends in big data analytics, Journal of Parallel and Distributed Computing, 74(7) (2014), pp.2561-2573.
[6] S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, on the use of mapreduce for imbalanced big data using random forest, Information Sciences, 285 (2014), pp.112-137.
[7] Z. Huang, A fast clustering algorithm to cluster very large categorical data sets in data mining, SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.
[8] O.Y. AL-Jarrah, P.D. Yoo, S.Muhaidat, G. K. Karagiannidis and K. Taha, Efficient machine learning for big data: A review, Big Data Research 2(3) (2015), pp 87-93.
[9] Changwon. Y, Luis. Ramirez and Juan. Liuzzi, Big data analysis using modern statistical and machine learning methods in medicine, International Neurourology Journal, 18 (2014), pp.50-57.
[10] P. Singh and B. Suri, 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 (2014), pp. 89-97.
[11] A. Jacobs, The pathologies of big data, Communications of the ACM, 52(8) (2009), pp.36-44.
[12]. H. Zhu, Z. Xu and Y. Huang, Research on the security technology of big data information, International Conference on Information Technology Management Innovation, 2015, pp.1041-1044.
[13]. Z. Hongjun, H. Wenning, H. Dengchao and M. Yuxing, Survey of research on information security in big data, Congress da sociedada Brasileira de Computacao, 2014, pp.1-6.
[14]. I. Merelli, H. Perez-sanchez, S. Gesing and D. Agostino, Managing, analyzing, and integrating big data in medical bioinformatics: open problems and future perspectives, BioMed Research International, 2014, (2014), pp.1-13.
[15] X. Y.Chen and Z. G.Jin, Research on key technology and applications for internet of things, Physics Procedia, 33, (2012), pp.561-566.
[16] M. D. Assuno, R. N. Calheiros, S. Bianchi, M. a. S. Netto and R. Buyya, Big data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, 79 (2015), pp.3-15.