

A Study of Research Issues & Challenges of Big Data Analytics

Neeraj Sharma^{1*} Dr. Mahaveer Sain²

¹ Research Scholar, Department of Computer, Science Maharishi Arvind University, Rajasthan

² Professor, Department of Computer Science & Informatics, Maharishi Arvind University, Rajasthan

Abstract – *An immense archive of terabytes of information is created every day from present day data frameworks and computerized innovations, for example, Internet of Things and distributed computing. Investigation of these monstrous information requires a great deal of endeavors at various levels to separate information for dynamic. Subsequently, enormous information examination is a momentum region of innovative work. The essential goal of this paper is to investigate the possible effect of huge information challenges, built-up examination issues, and different devices related with it. Thus, this article gives a stage to investigate large information at various stages. Also, it opens another skyline for scientists to build up the arrangement, in light of the difficulties and open examination issues.*

Keywords – Hadoop, Massive Data, Big Data, Structured Data, Unstructured Data, Big Data Analytics

1. INTRODUCTION TO BIG DATA

In advanced world, information are produced from different sources and the quick progress from computerized advances has prompted development of large information. It furnishes transformative discoveries in numerous fields with assortment of enormous datasets. As a rule, it alludes to the assortment of enormous and complex datasets which are hard to deal with utilizing customary information base administration instruments or information preparing applications. These are accessible in organized, semi-organized, and unstructured arrangement in petabytes and past. Officially, it is characterized from 3Vs to 4Vs. 3Vs alludes to volume, speed, and assortment. Volume alludes to the gigantic measure of information that are being produced ordinary though speed is the pace of development and how quick the information are assembled for being examination.

The Assortment gives data about the kinds of information, for example, organized, unstructured, semi-organized and so forth. The fourth V alludes to veracity that incorporates accessibility and responsibility. The prime goal of huge information investigation is to deal with information of high volume, speed, assortment, and veracity utilizing different conventional and computational savvy strategies.

A portion of these extraction techniques for getting accommodating data the accompanying Figure 1

alludes to the meaning of enormous information. Anyway careful definition for enormous information isn't characterized and there is an accept that it is issue explicit. This will help us in getting improved dynamic, understanding disclosure and enhancement while being imaginative and practical.

It is normal that the development of large information is assessed to arrive at 40 billion by 2018. From the point of view of the data and correspondence innovation, large information is a robust impulse to the up and coming age of data innovation ventures, which are extensively based on the third stage, basically alluding to huge information, distributed computing, web of things, and social business.

For the most part, Data distribution centers have been utilized to deal with the huge dataset. For this situation separating the exact information from the accessible huge information is a principal issue. The majority of the introduced approaches in information mining are not generally ready to deal with the huge datasets effectively. The key issue in the examination of huge information is the absence of coordination between data set frameworks just as with investigation devices, for example, information mining and measurable examination. These difficulties for the most part emerge when we wish to perform information revelation and representation for its viable applications.

Neeraj Sharma^{1*} Dr. Mahaveer Sain²

A key issue is the manner by which to quantitatively portray the fundamental qualities of large information. There is a requirement for epistemological ramifications in portraying information insurgency. Furthermore, the investigation on unpredictability hypothesis of large information will help comprehend basic attributes and arrangement of complex examples in enormous information, rearrange its portrayal, improves information reflection, and guide the plan of registering models and calculations on huge information. Much examination was completed by different scientists on large information and its patterns.

On the other hand, it is to be noticed that all information accessible as large information are not valuable for examination or dynamic cycle. Industry and the scholarly community are keen on scattering the discoveries of enormous information. This paper centers on difficulties in large information and its accessible strategies. Moreover, we state open examination issues in large information. Thus, to expand this, the paper is partitioned into following segments. Segments 2 arrangements with challenges that emerge during adjusting of large information. Area 3 outfits the open examination gives that will assist us with handling enormous information and concentrate helpful information from it. Area 4 gives an understanding to large information apparatuses and methods. End comments are given in segment 5 to sum up results.

2. CHALLENGES IN BIG DATA ANALYTICS

Getting on years enormous information has been amassed in a few areas like medical care, policy implementation, retail, biochemistry, and other interdisciplinary logical investigates. Online applications experience large information oftentimes, for example, social processing, web text and reports, and internet search ordering. Social registering incorporates social network investigation, online networks, recommender frameworks, notoriety frameworks, and expectation markets where as web search ordering incorporates ISI, IEEE Xplorer, Scopus, Thomson Reuters and so forth. Considering this favorable circumstances of large information it gives another open doors in the information preparing errands for the forthcoming analysts. Anyway opportunities consistently follow a few difficulties.

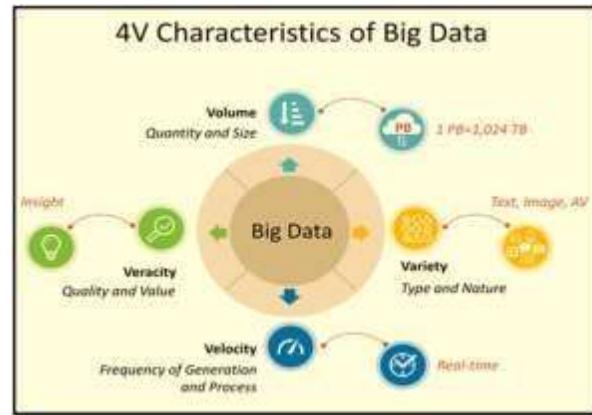


Figure 1: Showing Characteristics of Big Data

To deal with the difficulties we have to know different computational complexities, data security, and computational technique, to break down large information. For instance, numerous measurable techniques that perform well for little information size don't scale to voluminous information. Likewise, numerous computational strategies that perform well for little information face huge difficulties in breaking down enormous information.

The Different difficulties that the wellbeing segment face was being investigated by much scientists. Here the difficulties of large information investigation are arranged into four general classifications specifically information stockpiling and examination; information disclosure and computational complexities; adaptability and visualization of information; and data security. We examine these issues quickly in the accompanying subsections.

i. Data Storage and Analysis

Recently the size of information has developed exponentially by different methods, for example, cell phones, ethereal tangible innovations, distant detecting, radio recurrence distinguishing proof Perusers and so forth. These information are put away on spending a lot of cost while they disregarded or erased at last because there is no enough space to store them. Hence, the primary test for huge information examination is capacity mediums and higher info/yield speed. In such cases, the information openness must be on the first concern for the information revelation and portrayal.

The prime explanation is being that, it must be gotten too effectively and instantly for additional examination. In past decades, investigator utilize hard plate drives to store information however, it more slow arbitrary information/yield execution than successive information. To defeat this constraint, the idea of strong state drive (SSD) and expression change memory (PCM) was presented. Anyway the available stockpiling

advancements can't have the necessary exhibition for handling large information.

The test with Big Data investigation is credited to assorted variety of information with the ever developing of datasets; information mining undertakings has altogether expanded. Also information decrease, information choice, highlight determination is a fundamental errand particularly when managing enormous datasets. This presents a phenomenal test for analysts. It is because, existing calculations may not generally react in a sufficient time when managing these high dimensional information. Robotization of this cycle and growing new AI calculations to guarantee consistency is a significant test lately.

Notwithstanding all these Clustering of enormous datasets that help in examining the large information is of prime concern. Ongoing advances, for example, Hadoop and map Reduce make it conceivable to gather huge measure of semi organized and unstructured information in a sensible measure of time. The key building challenge is the means by which to viably examine these information for getting better information. A standard cycle to this end is to change the semi organized or unstructured information into organized information, and afterward apply information mining calculations to separate information.

The significant test for this situation is to give more consideration for structuring stockpiling systems and to hoist productive information examination apparatus that give ensures on the yield when the information originates from various sources. Besides, plan of AI calculations to break down information is basic for improving productivity and adaptability.

II. Knowledge Discovery and Computational Complexities

The Information disclosure and portrayal is a prime issue in enormous information. It incorporates various sub fields, for example, confirmation, documenting, the executives, safeguarding, information recovery, and portrayal. There are a few apparatuses for information disclosure and portrayal, for example, fluffy set, harsh set, delicate set, close to set, formal idea investigation, head part examination and so forth to give some examples.

Moreover many hybridized methods are likewise evolved to handle genuine issues. Every one of these strategies are issue subordinate. Further a portion of these procedures may not be reasonable for huge datasets in a consecutive PC. Simultaneously a portion of the methods has great qualities of adaptability over equal PC. Since the size of huge information continues expanding exponentially, the accessible instruments may not be proficient to deal with these information for acquiring important data. The most mainstream approach if

there should arise an occurrence of large dataset the board is information stockrooms and information shops. Information stockroom is primarily capable to store information that are sourced from operational frameworks though information shop depends on an information distribution center and encourages investigation.

The Investigation of enormous dataset requires more computational complexities. The significant issue is to deal with irregularities and vulnerability present in the datasets. As a rule, deliberate displaying of the computational multifaceted nature is utilized. It might be hard to build up an exhaustive numerical framework that is extensively pertinent to Big Data. However, a space explicit information investigation should be possible effectively by understanding the specific complexities. A progression of such improvement could reproduce enormous information investigation for various regions. Much examination and overview has been completed toward this path utilizing AI procedures with the least memory prerequisites. The essential target in these examinations is to limit computational cost preparing and complexities.

Nonetheless, current huge information examination devices have helpless performance in taking care of computational complexities, vulnerability, and irregularities. It prompts an extraordinary test to create methods and advances that can bargain computational complexity, uncertainty and irregularities in a viable way.

III. Scalability and Visualization of Data

The most significant test for huge information examination techniques is its versatility and security. In the most recent decades specialists have paid considerations to quicken information examination and it's accelerate processors adhered to by Moore's Law. For the previous, it is important to create testing, on-line, and multiresolution examination methods. Steady strategies have great versatility property in the part of huge information investigation.

As the information size is scaling a lot quicker than CPU speeds, there is a characteristic emotional move in processor innovation being inserted with expanding number of centers. This move in processors prompts the advancement of equal registering. Continuous applications like route, interpersonal organizations, fund, web search, idealness and so on requires equal processing.

The goal of imagining information is to introduce them all the more satisfactorily utilizing a few procedures of diagram hypothesis. Graphical representation furnishes the connection between information with appropriate interpretation. Nonetheless, online commercial center like Flipkart, amazon, e-cove have a large number of

clients and billions of products to sold every month. This produces a great deal of information. To this end, some organization utilizes a device Tableau for enormous information perception. It has capacity to change huge and complex information into instinctive pictures. This assistance representative of an organization to picture search pertinence, screen most recent client feedback, and their slant examination. Notwithstanding, current enormous information representation apparatuses generally have lackluster showings in functionalities, adaptability, and reaction in time.

We can see that large information have delivered numerous challenges for the advancements of the equipment and programming which prompts equal figuring, distributed computing, and distributed registering, perception measure, versatility. To overcome this issue, we have to correspond more numerical models to software engineering.

IV. Information Security

In large information examination monstrous measure of information are associated, investigated, and dug for significant examples. All associations have various arrangements to safe watchman their delicate data. Saving touchy data is a significant issue in large information investigation. There is a gigantic security hazard related with large information.

Accordingly, data security is turning into a major information investigation issue. Security of large information can be upgraded by utilizing the methods of validation, approval, and encryption. Different safety efforts that large information applications face are size of system, wide range of gadgets, constant security observing, and absence of interruption framework. The security challenge brought about by enormous information has pulled in the consideration of data security. Thusly, consideration must be given to build up a staggered security strategy model and counteraction framework.

3. RESEARCH ISSUES IN BIG DATA ANALYTICS

The Big data analytics and data science are turning into the exploration point of convergence in ventures and the scholarly world. Information science targets investigating enormous information and information extraction from information. Uses of large information and information science incorporate information science, vulnerability demonstrating, dubious information examination, AI, measurable learning, design acknowledgment, information warehousing, and sign preparing.

The Viable mix of advancements and investigation will bring about anticipating the future float of occasions. Primary focal point of this area is to examine open examination issues in enormous

information investigation. The exploration issues relating to large information investigation are arranged into three general classes to be specific web of things (IoT), distributed computing, bio motivated registering, and quantum figuring. Anyway it isn't restricted to these issues.

I. IOT for Big Data Analytics

The Internet has rebuilt worldwide interrelations, the specialty of organizations, social transformations and a mind blowing number of individual attributes. At present, machines are getting in on the demonstration to control incalculable independent devices through web and make Internet of Things (IoT). Consequently, apparatuses are turning into the client of the web, much the same as people with the internet browsers. Web of Things is drawing in the consideration of late scientists for its most encouraging chances and difficulties. It has a basic financial and cultural effect for the future development of data, system and correspondence innovation. The new guideline of future will be in the end, all that will be associated and wisely controlled.

The idea of IoT is getting more appropriate to the practical world because of the advancement of portable devices, installed and universal correspondence advances, distributed computing, and information investigation.

Besides, IoT presents difficulties in mixes of volume, speed and assortment. From a more extensive perspective, much the same as the web, Internet of Things empowers the gadgets to exist in a bunch of spots and encourages applications going from unimportant to the vital. Then again, it is as yet beguiling to comprehend IoT well, including definitions, substance and contrasts from other comparative ideas. A few broadened innovations, for example, computational insight, and huge information can be joined together to improve the information the executives and information revelation of enormous scope mechanization applications.

The Information obtaining from IoT information is the greatest test that huge information proficient are confronting. In this manner, it is fundamental to create framework to investigate the IoT information. An IoT gadget creates ceaseless surges of information and the scientists can create instruments to separate important data from these information utilizing AI strategies.

To Understanding these surges of information produced from IOT gadgets and dissecting them to get significant data is a difficult issue and it prompts huge information examination. AI calculations and computational insight a strategy is the just answer for handle huge information from IoT forthcoming. Key innovations that are

related with IOT are additionally talked about in many examination papers. Figure 2 delineates a diagram of IOT enormous information and information revelation measure.

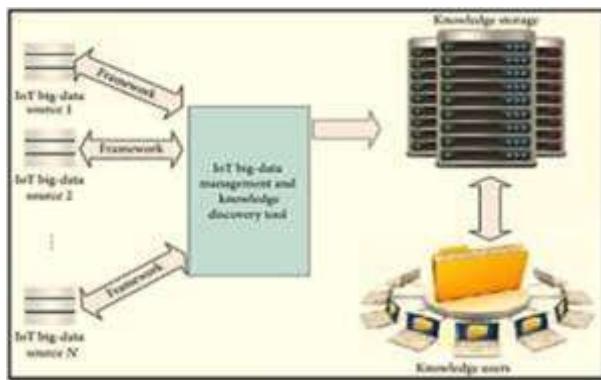


Figure 2: Showing IOT Big Data Knowledge Discovery

The Information investigation framework have begun from hypotheses of human data handling, for example, outlines, rules, labeling, and semantic systems. By and large, it comprises of four fragments, for example, information procurement, information base, information dispersal, and information application.

In information procurement stage, information is found by utilizing different conventional and computational knowledge methods. The found information is put away in information bases and master frameworks are commonly planned dependent on the found information. Information dispersal is significant for acquiring important data from the information base. Information extraction is a cycle that searches archives, information inside records just as information bases. The last stage is to apply found information in different applications. It is a definitive objective of information disclosure.

The information investigation framework is essentially iterative with the judgment of information application. There are numerous issues, conversations, and investigates in this subject matter investigation. It is past extent of this review paper. For better perception, information investigation framework is portrayed in Figure 3.

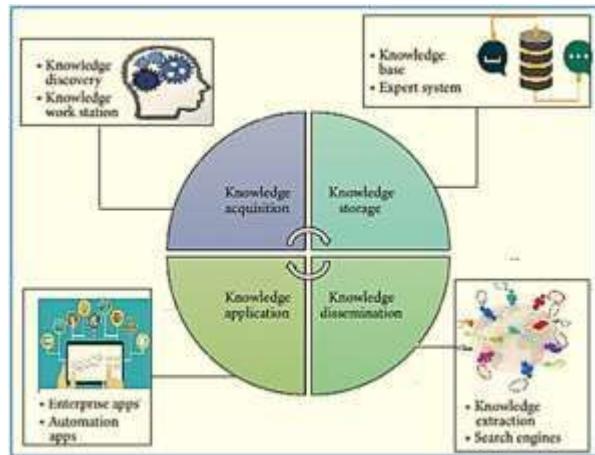


Figure 3: Showing IOT Knowledge Exploration System

4. TOOLS FOR BIG DATA PROCESSING

The Large quantities of instruments are accessible to deal with enormous information. In this area, we examine some current procedures for dissecting large information with accentuation on three significant developing apparatuses in particular Map Reduce, Apache Spark, and Storm. The majority of the accessible instruments focus on clump handling, stream processing, and intuitive investigation. Most clump preparing instruments depend on the Apache Hadoop framework, for example, Mahout and Dryad. Stream information applications are generally utilized for ongoing logical. A few instances of huge scope streaming stage are Strom and Splunk. The intelligent examination measure permits clients to legitimately communicate progressively for their own investigation.

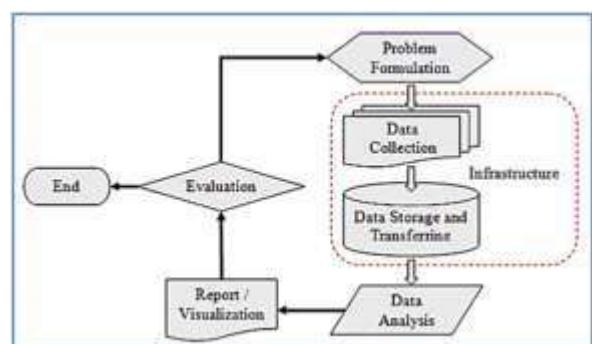


Figure 4: Showing Big Data Projects Workflow

I. Apache Hadoop and Map Reduce

The most settled programming stage for enormous information examination is Apache Hadoop and Map reduce. It comprises of Hadoop part, map reduce, Hadoop disseminated document framework (HDFS) and apache hive and so forth. Guide lessen is a programming model for preparing

enormous datasets depends on partition and vanquish technique.

The gap and overcome technique is executed in two stages, for example, Map step and Reduce Step. Hadoop chips away at two sorts of hubs, for example, ace hub and specialist hub. The ace hub isolates the contribution to littler sub issues and afterward conveys them to specialist hubs in map step. From that point the ace hub joins the yields for all the sub problems in decrease step. In addition, Hadoop and Map Reduce fill in as a ground-breaking programming structure for tackling huge information issues. It is additionally useful in issue open minded stockpiling and high throughput information preparing.

II. Apache Mahout

Apache mahout means to give adaptable and business AI strategies for huge scope and smart information investigation applications. Center calculations of mahout including bunching, arrangement, design mining, relapse, dimensionality decrease, transformative calculations, and cluster put together shared sifting run with respect to head of Hadoop stage through guide diminish structure. The objective of mahout is to fabricate a dynamic, responsive, assorted network to encourage conversations on the undertaking and potential use cases.

The fundamental goal of Apache mahout is to give an instrument to alleviating huge difficulties. The various organizations the individuals who have actualized versatile AI calculations are Google, IBM, Amazon, Yahoo, Twitter, and Facebook.

III. Apache Spark

Apache spark is an open source enormous information preparing system worked for speed handling, and complex investigation. It is anything but difficult to utilize and was initially evolved in 2009 in UC Berkeleys AMP Lab. It was publicly released in 2010 as an Apache venture. Flash lets you rapidly compose applications in java, scala, or then again python. Notwithstanding map decrease tasks, it upholds SQL questions, streaming information, AI, and diagram information handling. Flash sudden spikes in demand for head of existing Hadoop appropriated record framework (HDFS) foundation to give upgraded and extra usefulness. Flash comprises of parts to be specific driver program, group director and specialist hubs.

The driver program fills in as the beginning stage of execution of an application on the sparkle bunch. The group supervisor assigns the assets and the laborer hubs to do the information preparing as undertakings. Every application will have a lot of cycles considered agents that are liable for executing the assignments. The significant bit of leeway is that

it offers help for sending flash applications in a current Hadoop bunches.

The Figure 5 depicts the design outline of Apache Spark. The different highlights of Apache Spark are recorded beneath:

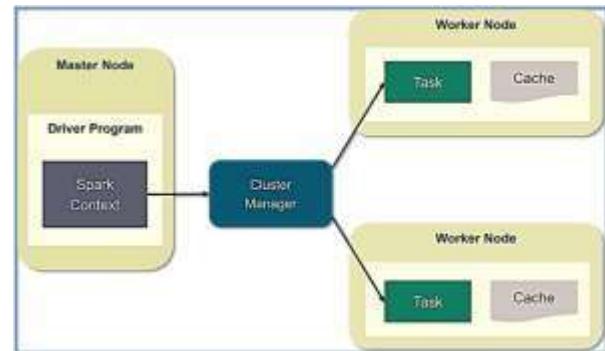


Figure 5: Showing Architecture of Apache Spark

- The prime focal point of sparkle incorporates tough conveyed datasets (RDD), which store information in memory and give adaptation to non-critical failure without replication. It underpins iterative calculation, improves speed and asset use.
- The preeminent bit of leeway is that notwithstanding Map Reduce, it additionally bolsters streaming information, AI, and chart calculations.
- Another bit of leeway is that, a client can run the application program in various dialects, for example, Java, R, Python, or Scala. This is conceivable as it accompanies more elevated level libraries for cutting edge examination. These standard libraries increment designer efficiency and can be consistently consolidated to make complex work processes.
- Spark assists with running an application in Hadoop bunch, up to multiple times quicker in memory, and multiple times quicker when running on plate. It is conceivable as a result of the decrease in number of peruse or compose activities to circle.
- It is written in Scala programming language and runs on java virtual machine (JVM) condition. Furthermore, it supports java, python and R for creating applications utilizing Spark.

IV. 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 use 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 [37].

V. Storm

The Storm is a dispersed and flaw lenient ongoing calculation framework for preparing huge streaming information. It is uniquely intended for ongoing preparing in appears differently in relation to Hadoop which is for clump handling. Furthermore, it is additionally simple to set up and work, adaptable, flaw lenient to give serious exhibitions. The tempest bunch is obviously like Hadoop group.

On storm bunch clients run various geographies for various tempest undertakings while Hadoop stage executes map diminish occupations for comparing applications. There are number of contrasts between maps lessen occupations and geographies.

The essential contrast is that guide lessen work in the long run completes though a geography measures messages constantly, or until client end it. A tempest bunch comprises of two sorts of hubs, for example, ace hub and laborer hub. The ace hub and laborer hub actualize two sorts of jobs, for example, aura and manager separately. The two jobs have comparable capacities as per job tracker and task tracker of guide lessen system. Glow is accountable for appropriating code over the tempest group, booking and allotting assignments to laborer hubs, and checking the entire framework.

The manager consents undertakings as allocated to them by glow. Likewise, it begin and end the cycle as important dependent on the directions of glow. The entire computational innovation is divided and appropriated to various specialist measures and every laborer cycle executes an aspect of the topography.

VI. Apache Drill

The Apache drill is another distributed system for interactive analysis of big data. It has more flexibility to support many types of query languages, data formats, and data sources. It is also specially designed to exploit nested data. Also it has an objective to scale up on 10,000 servers or more and reaches the capability to process petabytes of data and trillions of records in seconds. Drill use HDFS for storage and map reduce to perform batch analysis.

VII. Jaspersoft

The Jaspersoft package is an open source software that produce reports from database columns. It is a scalable big data analytical platform and has a capability of fast data visualization on popular storage platforms, including Mongo DB, Cassandra, Redis etc. One important property of Jasper soft is that it can quickly explore big data without extraction, transformation, and loading (ETL). In addition to this, it also have an ability to build powerful hypertext markup language (HTML) reports and dashboards interactively and directly from big data store without ETL requirement. These generated reports can be shared with anyone inside or outside user's organization.

VIII. Splunk

In recent years a lot of data are generated through machine from business industries. Splunk is a real-time and intelligent platform developed for exploiting machine generated big data. It combines the up-to-the-moment cloud technologies and big data. In turn it helps user to search, monitor, and analyze their machine generated data through web interface. The results are exhibited in an intuitive way such as graphs, reports, and alerts.

The Splunk is different from other stream processing tools. Its peculiarities include indexing structured, unstructured machine generated data, real-time searching, reporting analytical results, and dashboards. The most important objective of Splunk is to provide metrics for many application, diagnose problems for system and information technology infrastructures, and intelligent support for business operations.

5. SUGGESTIONS FOR FUTURE WORK

The measure of information gathered from different applications everywhere on over the world over a wide assortment of fields today is relied upon to twofold like clockwork. It has no utility except if these are examined to get valuable data. This requires the advancement of methods which can

be utilized to encourage huge information investigation.

The improvement of incredible PCs is a help to actualize these procedures prompting mechanized frameworks. The change of information into information is in no way, shape or form a simple assignment for elite enormous scope information preparing, including abusing parallelism of current and forthcoming PC designs for information mining. Also, these information may include vulnerability in various structures. A wide range of models like fluffy sets, harsh sets, delicate sets, neural systems, their speculations and half and half models got by consolidating at least two of these models have been discovered to be productive in speaking to information. These models are additionally especially productive for examination.

Usually, large information are decreased to incorporate just the significant qualities essential from a specific report perspective or depending upon the application region. Thus, decrease methods have been created. Regularly the information gathered have missing qualities. These qualities should be produced or the tuples having these missing values are wiped out from the informational index before examination. All the more critically, these new difficulties may contain, here and there even disintegrate, the exhibition, effectiveness and adaptability of the devoted information concentrated registering frameworks.

The later methodology at times prompts loss of data and consequently not liked. This raises many exploration issues in the business and examination network in types of catching and getting to information viably. What's more, quick handling while at the same time accomplishing superior and high throughput, and putting away it productively for some time later is another issue. Further, programming for large information investigation is a significant testing issue. Communicating information access necessities of uses and structuring programming language deliberations to abuse parallelism are a prompt need.

Moreover, AI ideas and devices are picking up prominence among scientists to encourage significant outcomes from these ideas. Analysis in the region of machine learning for huge information has zeroed in on information preparing, calculation execution, and streamlining. A significant number of the AI apparatuses for large information are begun as of late needs exceptional change to embrace it. We contend that while every one of the instruments has their points of interest and constraints, more productive apparatuses can be created for managing issues inalienable to huge information. The proficient apparatuses to be created must have arrangement to deal with uproarious and awkwardness information, vulnerability and irregularity, and missing qualities.

6. CONCLUSIONS

In recent years data are generated at a dramatic pace. Analyzing these data is challenging for a general 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.

REFERENCES

1. A. Nugent, F. Halper, and M. Kaufman (2018). —Big Data for Dummiesll: John Wiley & Sons.
2. A. Gandomi and M. Haider (2015). Beyond the hype: —Big data concepts, methods, and analyticsll, International Journal of Information Management, 35(2), pp. 137-144.
3. D. Che, M. Safran, and Z. Peng (2018). "From Big Data to Big Data Mining: Challenges, Issues, and Opportunities," in Database Systems for Advanced Applications, pp. 1-12.
4. M. K.Kakhani, S. Kakhani and S. R.Biradar (2015). —Research issues in big data analyticsll, International Journal of Application or Innovation in Engineering & Management, 2(8), pp. 228-232.
5. X. Jin, B. W.Wah, X. Cheng and Y. Wang (2018). —Significance and challenges of big data researchll, Big Data Research, 2(5), pp. 58-63.
6. C. L. Philip, Q. Chen and C. Y. Zhang (2018). —Data-intensive applications, challenges, techniques and technologies: A survey on big data, Information Sciences, 278, pp. 312-345.
7. K. Kambatla, G. Kollias, V. Kumar and A. Gram (2014). Trends in big data analyticsll, Journal of Parallel and Distributed Computing, 74(7), pp. 2561-2573.

8. MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell (2014). Health big data analytics: current perspectives, challenges and potential solutions, International Journal of Big Data Intelligence, 1, pp. 114-126.
9. R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese (2017). A look at challenges and opportunities of big data analytics in healthcare, IEEE International Conference on Big Data, pp. 15-20.
10. R. Jain (2018). "Big Data Fundamentals".
11. T. H. Davenport, P. Barth, and R. Bean (2017). "How Big Data is different," MIT Sloan Management Review, vol. 54.
12. N. Mishra, C. Lin and H. Chang (2015). —A cognitive adopted framework for iot big data management and knowledge discovery prospectivell, International Journal of Distributed Sensor Networks, pp. 1-13
13. Z. Huang (2007). A fast clustering algorithm to cluster very large categorical data sets in data miningll, SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery.

Corresponding Author

Neeraj Sharma*

Research Scholar, Department of Computer, Science Maharishi Arvind University, Rajasthan

nesh787@rediffmail.com