

Challenges and Issues of Data Analytics in Emerging Scenarios for Big data, Cloud and Image Mining

A Madhuri¹, S. Phani Praveen², D Lokesh Sai Kumar³, S Sindhura⁴, Sai Srinivas Vellela⁵

^{1,3}Assistant Professor, Department of Computer Science and Engineering, PVPSIT,
Vijayawada, A.P, India.

⁴Assistant Professor, Department of Computer Science and Engineering, KoneruLakshmaiah
Education Foundation, Guntur, A.P, India.

⁵Assistant Professor Department of Computer Science and Engineering, Chalapathi Institute
of Technology, Guntur, AP, India.

madhuri.abburi@gmail.com¹, phani.0713@gmail.com², lokeshsaikumar@gmail.com³,
ssindhurapraveen@gmail.com⁴, sais1916@gmail.com⁵

Abstract: During the digital age, business leaders had vast amounts of data accessible. Significant knowledge is referred to as databases not only comprehensive but also broad in size and speed, which renders conventional methods and techniques challenging to use. Solutions must be explored and supplied so that these datasets can manage and derive meaningful information due to the rapid development of such data. Decision-makers will, therefore, be willing, through routine activities to consumer communications and social network data, to extract useful knowledge from these diverse and quickly evolving data. This can be done by implementing the latest statistical methods on Large Data using extensive data analytics. Built-in, distributed, distributed, fault-tolerant, flexible and accessible architectures are being widely used in cloud environments for massive computational applications. The HDFS architecture is planned to identify faults, such as accidents of call-nodes, built-in node failures, and network failures, and route, built-in to further integrate processes. Redundancy offers an essential location for facts when running on large sets of information. A backup procedure ensures that the data is available and accessible. A big intuition challenge occupies the bulk of the insufficiency contained inside the current results. The determination of this paper is to explore different analytical methods and tools which can be applied to Big Data and the benefits given by the use of Big Data Analytics in several judgements.

Keywords: *Data analysis, Big data, Cloud environment, Computational challenges, Image mining, scalability challenges.*

1. Introduction:

Data analytics is used for informational science, math, statistics, and other disciplines; this research's nature is uniquely mysterious: challenging scientific questions and pressing social problems. Data Analysis of its volume, speed, grouping, and variance, gigantic data overwhelms current institutions and programming throughout this way, it is essential to understand each of these features when considering the specifications for complete, monumental data analysis applications. Some elements are critical to promote a comprehensive data analysis and to address the differences within applications that are sensitive to pause and development[1]. Acquisition: all data sources and any arrangements

and structures envisaged must be recognized. To think about the handling capabilities, the data needed must be portrayed and recorded. Furthermore, the amount of data anchorage is vital to consider, as they cope, and restriction of plans is affected. The time allotted to the data assembly and age is a further important point. Some knowledge is generated automatically, while other details are driven on and typically analysed when they arrive.

Defending: The use, incarceration, transparency, and constancy of capable restricted and association instruments are used to increase storage capacity. A few data from different sources are accessible and are not affected by computer developers; in any event, other data is collected via the program [2]. Problems of teamwork and comparability would preferably be timely. Furthermore, the amount of information must increase sensibly after a period, and the storage space and instrument used are taken care of. This will help determine whether to buy additional hardware for keeping all data internally, for utilizing current cloud systems, or for using a combination of them. Application fashioners should consider the data's general characteristics necessary for the request and should have specific images of how they are handled and secured [3]. Those requirements must be evaluated and unquestionably documented for the client to better address requirements.

Reprocessing: Most data collected can already be processed to obtain partial results, which can then help the correct planning to achieve the results required. Possible processing techniques combine to orchestrate the views of other fields, need various keys, particular compact goals, and add markers and metadata to unstructured data formats. The pre-processing methods combine.

Installation: An program can utilize transparent data with great performance. Regardless of the gigantic scale of data sets, the review is a magnificent, full, and exhaustive procedure that requires to be changed to better conduct the necessities that requirements of the program.

We discuss in this paper in various sections about Challenges and issues that's prescriptive analysis, arrangement of challenges, massive computational challenges, and Image mining and scalability challenges.

2. Prescriptive Analytics

An emerging method of analytics is called prescriptive analytics, which proposes one or more publications of the campaign and illustrates the possible outcomes of each decision. The first move is to perform analyses of specific knowledge to satisfy the requirement for commercial businesses to interpret details in multiple outlets such as similar databases, Excel files, Twitter and Facebook. This paper should tackle the challenging circumstances to allow predictive big statistical analytics. The requests for specific, half-dependent and unstructured current codecs are spread through numerous knowledge centres in connection repositories, NoSQL repositories and file systems. They have to be put based on fact-mining algorithms into a format [4]. Most existing libraries are using extract-remodel loading to extract records from single stores and to restructure their layout into a corresponding schema. It takes time, and all the facts must be acquired in advance.

2.1 Consumable Massive Facts Analysis

The consumable massive facts analysis was also to be taken care, It makes it difficult for companies to obtain relevant specialised training to undertake huge consumable examination because the research is multidimensional. The consumable search presents essential features for managing this task and for overcoming the inaccessibility of analytical capacities by means of reducing the application of the survey [5]. Consumer analysis refers to the development of capabilities that are efficient and existing in business by means of creating equipment that facilitates the creation, over-seeing and spending of an enquiry. The commodity inquiry is a public platform or a dialect for the management of knowledge regarding human resources such as circulatory pressure, weight and the amount of sugar.

2.2 The Mining Algorithm Challenge

The mining algorithms also plays vital role, Most individuals with current library services such as R, WEKA, and RapidMiner only support the simultaneous implementation of facts-mining algorithms by one-gadget. This tends to make the above libraries wrong to handle large volumes of vital records. In decentralized optimized extraction libraries such as Apache Mahout, Cloudera Oryx, Oxdata H2O, MLlib and Deep learning, records of Hadoop and Spark mining algorithms are re-written. These libraries are progressed by searching for and rebuilding algorithms to parallel modules. This is a complex, time-consuming process and the happily updated collection of rules is solely focused on input informants. This makes it difficult to expand, maintain and expand these libraries and includes massive statistics particularly which are important [6]. It is necessary to evaluate the evidence from which you rely upon. In accordance with McKinsey's examination of extensive data as massive facts, a further need for IT experts is a commitment to large databases: the next borderline for modern architecture. These documents are evidence that a corporation must either hire professionals or train existing staff in the brand spanking new profession to take the vast records strategy.

2.3 Hardware Challenge

This influences the mathematical storage system and makes it extremely difficult to plaster; it can establish an endless linking among the devices that deliver information to the network. The "sender" means that there is no difference between the "receiver" and the information to be processed. This loop should be painted while the data received by the framework tells the device to avoid the facts that are only sent. Would you miss evidence for a clear assessment method that will save you? Throughout the implementation procedure, this method can be lent. To prevent this, senders will create a "key" for any information that is transmitted. This solution is comparable to the MD5 Hash produced by compressed content materials to boost expertise. But the keys are robotically compared in this example. Record failure is not a chronic concern for hardware. The software can also work badly and cause irreversible and more risky risk of data loss. If one hard drive fails, there's usually another that can back it up, and knowledge doesn't suffer. However, owing to the computer "bug" or the configuration error of software failures, the data are always lost. Programmers have built a variety of methods to solve this problem, raising the effect of a malfunction in software. An interesting

case is Microsoft Word, which sometimes preserves the pictures a customer produces to shield against their lost opportunity in the case of a hardware or software system.

2.4 Integration Challenge

In this stage, data is aggregated and reworked in the proper layout for further statistical assessment. The primary challenge of Big Data Analytics (BDA) is to integrate unstructured data. Irrespective of the integration of data from the formed electronic health record (EHR), as illustrated in figure. 1 various difficulties occur in the transmission to the MySQL database of device Y when the fixed health information saved to an Oracle database in device X is. To hold records, Oracle and MySQL servers use computational constructs. In comparison, Program X may use the type of details "quantity" to store attractiveness statistics for patients, while Device Y may use the kind of knowledge "CHAR." Recorded metadata describes a resource's properties. The column names are used as metadata to define the features of stored statistics in the relational database format.

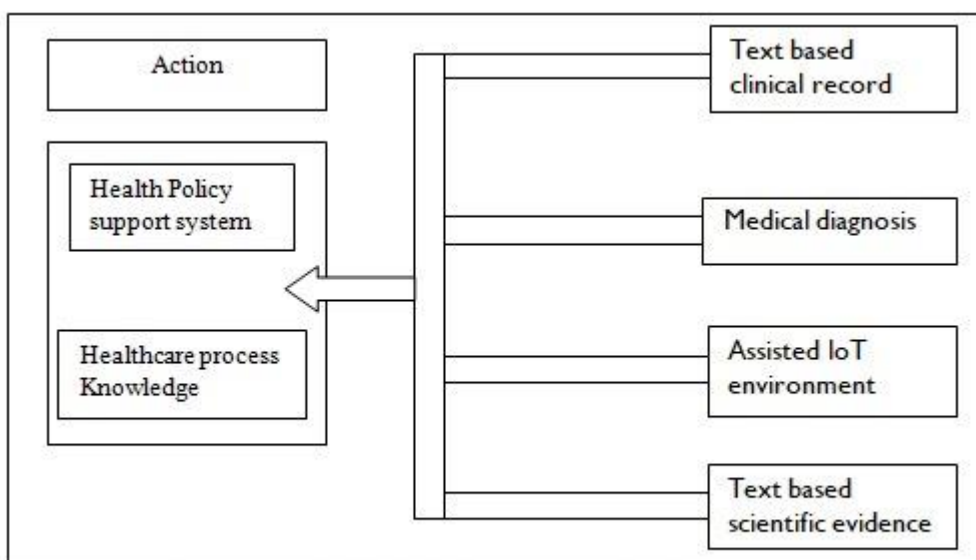


Figure: 1. Electronic Health Care Record Integration

Two major problems exist in the integration of metadata. To explain the material, different database systems first use single metadata. For instance, one method may use "sex" when someone talks about "gender" at the same time. A PC doesn't know the semantic relation between "sex" and "gender." Second, when mapping simple metadata to composite metadata, there are problems. In composite metadata "FirstName" + "LastName" inside the device, for instance, a PC cannot robotically map a PatientName "metadata into the system. Besides, code map problems could be addressed by various arrangements using one-of-a-kind relevant data analysing strategies. For example, the SNOMED-CT and ICD-10 codes for the "foot abscess" disease are unique. There is no coding device for moving to the map.

2.5 Pattern Interpretation Challenge

Besides, many believe that absolute clarity provides better data for estimation regularly. The devices of great news and know-how will not protect us against skews, gaps, and insufficient false assumptions. Be that as it may, some other attempt demonstrates that for massive data collections, substantial costs are standard if the goal is to ensure the information is as transparent as real-world information appears in Figure. 2.

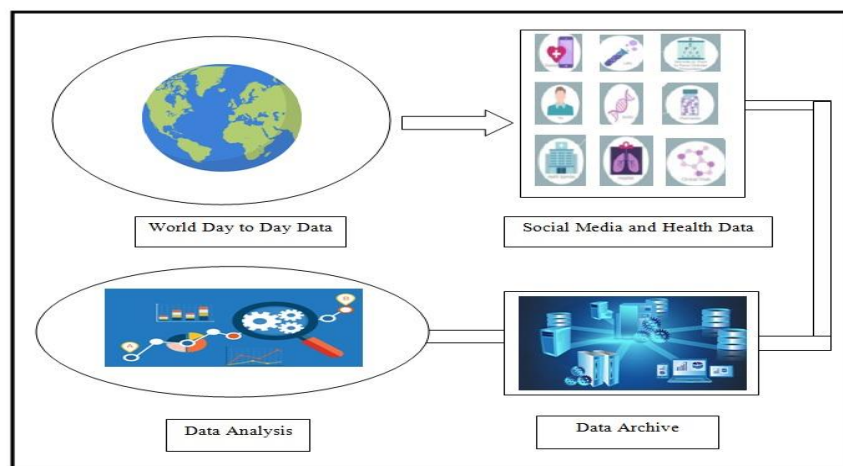


Figure: 2. Data Archive and Analysis

3. Arrangement of Challenges

Unstructured insights can be integrated and systematized in an uncooked fashion very difficultly. In this storage, records are performed to distinguish essential and sensitive information from fresh ideas. Besides, some answers for the non-structured reconciliation of data have been proposed. The problem with these methodologies is that a large number of them are uncomfortably situated, which is to say that the method is most effectively updated to examine evidence indexes one by one [7]. There are not many non-exclusive systems to attach unstructured data. Responses are arranged into fundamental techniques to subordinate data establishment.

3.1 User Intervention Method

Continually makes errors for computerized construction (a component of metadata) or case-matching metric measurements. Several territorial experts can consistently make these mistakes. However, this technique is impractical for large-scale data joining since it requires too much metadata to manually check the faults of multiple researchers, using swarm remark to update the embedded. A decision-based approach that relates to the complicated problems of knowledge inclusion involves a case of this. A PC architecture initially supplied the program with hundreds of practical social security mappings. After that, it accepted by way of organizing rules the most severe, logical equivalent for tables and related disciplines [8]. These regulations protect the various forms of the semantic of public resources systems. This pattern is first obtained by the dramatic schematic mapping precision of client improvisations.

3.2 The Probabilistic Method

The probabilistic integration system offers possibilities for relations between sets of pattern instruments. Once you evaluate the probabilities, a limit is applied to choose things that do not match. The insecurity generated during the mixing process is thus removed. The probabilistic methodology aims to mechanically generate an involved structure from agreed knowledge sources and the apparent textual maps between the properties and the cross-construction. It is not concerned about human intercession. Also, what method should be tested before they are sent? Different methodologies approve the styles: [1] Provides observable validity in order to determine if the documents or model have problems; [2] isolates data in the planning and test sets to verify design accuracy, [3] needs space experts to investigate what the structures found are, which includes the examination of the focal condition.

We use privacy saving details that decide the estimates for learning disclosures to maintain the secrecy of safety and handle security-required circumstances. Governments should also create robust laws to ensure the confidentiality of records [9]. Difficulties in offering public guarantees have been established. Extensive data are presently conveyed in order to create possible future policy benefits, even if the user can generate more challenges, e.g., administrative scenarios, top-of-the-line information, and issues concerning privacy, which all can be achieved by using extensive information in a total population procedure in an area..

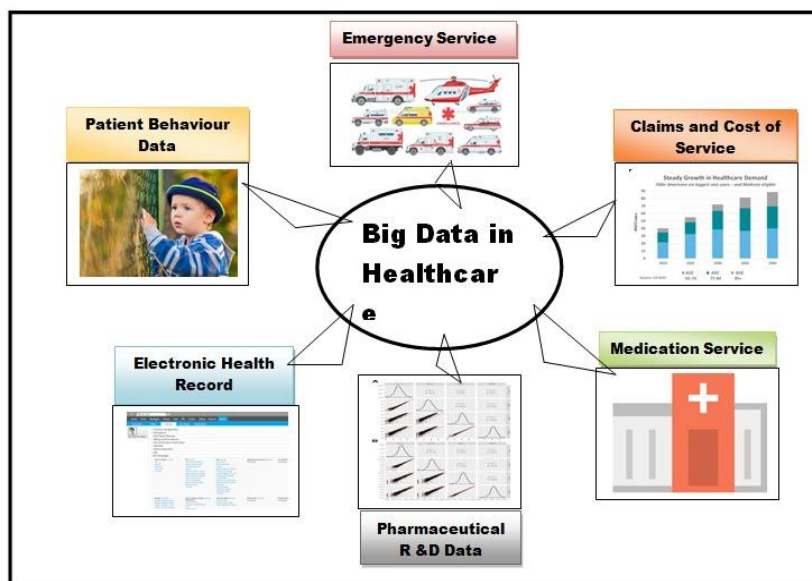


Figure: 3. Human body as a Source of Data

3.3 The challenge of Detecting Anomalies

The general classification of abnormality is an endeavour that aims to change a degree or identify anomalies in human environments. The existence and identification of the social "abnormalities" and the variations within them are less evident. In contrast, abnormalities are found in the field of disease outbreaks or malfunctions in certain kinds of complex systems, such as improved automobile engines [10]. The scale of data grows rising exponentially every day, and a great deal of health data is being generated as Figure 3 indicates.

4. Massive Computation Challenges

4.1 HADOOP

Hadoop tools are the first step to evolve to vast quantities of established, semi-structured and unstructured data. Many experts are fascinated by Hadoop as another breakthrough. Much of the sources must be learnt, and the eye is at some point shifted from the establishment of the primary goal towards Hadoop. Apache Hadoop is an open-source version of Google's existing Mapreduce framework. It enables the continuous treatment of petabytes to request data sets across hundreds or thousands of based product PCs. Parallel systems have been used regularly to use a wide range of data during a test [11]. Hadoop's two main fragments are shown in the two accompanying sections: HDFS and Mapreduce.

4.2 Hadoop Distributed File System

The Hadoop-Distributed File System (HDFS) is the only portion of Hadoop; generous illuminating accumulations are expected to be stored regularly in clusters and streamed to customers with high-performance applications. HDFS efficiently stores metadata of the record structure and program files. It generally saves three freely replicated copies of each datum square to ensure genuine quality, transparency and performance.

4.3 Mapreduce Hadoop

Hadoop Mapreduce is a distributed design parallel programming model executed over HDFS. A Job Tracker and a few Task Trackers are in the Hadoop Mapreduce motor. The JobTracker divides it in smaller errands (outline reduction) man-aged by the Task Tracker right when a Mapreduce task is executed. In the map step, the centre point divides the information into smaller sub-problems and transfers it to workers' centres. Increasing centre point forms a sub-component and outputs its effects as a core. The features with the corresponding key are combined in the reduction stage and organized by a related system to show the last output.

4.4 Apache Spark

Apache Spark is an open-source in-memory data processing program built in the UC Berkeley AMPLab for figurative structure. Spark also has impressive functionality such as mobility and internal manipulation, as Mapreduce does, like a Mapreduce like collecting and enrolled device. Spark is impressed by the Resilient Distributed Data Sets (RDDs), which make Spark a complete program which fits the requirements of iterative companies such as PageRank calculations, K-suggestions, and so on. RDD is eligible to Ignite and independent Start from Mapreduce conventional motors as well. Besides, given RDDs, Spark implementations can retain the data in a memory that is perishing in the middle of dissatisfaction by requesting and reproducing such data. RDD is a scanned array of details that can either be stored in an actual boundary system (e.g. HDFS) or can be caused by multiple RDDs. RDDs store loads of knowledge, including its dissemination and behaviour on parent RDDs known as a legacy, and Spark retrieves missing data quickly and efficiently. It begins to show high performance in preparing iterative estimates since it can reuse direct results and store data through several parallel commitments.

5. Image Mining with big data Challenges

Various therapeutic approaches were proposed to divide the image, and numerous vital inventions were acquired. However, medicinal photographs can include distinctive assortments of ancient rarities because of shortcomings in social security imaging systems. These rare old things can affect the information on the subject and stupide the pathology. The enticing imaging advancement can moderate some artifacts, and others require corresponding supervision [12]. Natural marvels generated in clinical science are disruptions, no uniformity of force, and incomplete volume impacts deemed the outstanding problems in a therapeutic division. An image can be positioned in areas that could be homogeneous by different processes. Given the multi-faceted consistency and mistakes, not all approaches are suitable for medical research. No typical picture division method will deliver adequate results for all imaging applications such as mind RMI, brain development research, etc. The ideal description of highlights, muscles, brain, and non-mental elements is known as essential constraints for separation of the mind. Another interference is the same division into an excess of the full field of view [13]. The bearings and operational levels of managers are often conventional boundaries to cognitive heterogeneity. Another reason for the problem lies in the division treatment.

The question of the expulsion of closer views from the foundation in the picture will be the distribution of images. This is perhaps the most critical issue of PC vision, and it has fascinated many investigators continuously. Because PCs are being used widely and progressively over time, more applications in new, restorative, and person fields need accurate picture splitting. Owing to the wide variety of possible objects, a fully structured division remains available, rendering "clues" among human beings inevitable. Intelligent picture division is becoming more and more widespread among seekers. The intuitive division's aim is (a) exact usage of consumer details in a way that needs negligible conversation and insubstantial response times by separating the object(s) from the base. This proposal would undoubtedly continue with the introduction of general methods of sorting division strategies, a detailed analysis of current type-dependent original methods of picture division, and conclude with the presentation of a modern dramatically modified modification and division technique [14]. The separation of pictures is the main problem of image analysis and photo comprehension. This is also an essential issue of PC vision and illustration identification [15]. Active contour models (ACM) are the best picture division processes, and the critical idea of ACM is to pursue a bend as defined by certain constraints to separate the protest required. Those standard dynamic form models classified as edge-based and based are two kinds that are characterized by their special paybacks and negatives and control the decision in application areas with the different characteristics of the images [16][17]. The model forms an edge-based capacity, which can generate the shape at the boundaries of the protests. With exceptional clamour or possibly a small edge, the edge-based dimension by the picture-next can investigate the correct confinements for photos.

6. Scalability Challenges

6.1 Big Data and Cloud Computing

In comparison to an on-site system, the cloud's reaction is rendered in a less complicated and more natural manner. Holding recent affairs locked within a dataset will eliminate confusion because the system can be grounded and the environment divided such that it can be nearly infinite. The network-connected storage contains a case of in-house accumulation of extensive data. The design will continue with a NAS case with many PCs attached to a Computer used as a NAS unit. The setup would be performed with a NAS program. A few NAS units will be linked through the CPU used as the NAS contraction. Stocking NAS collected is costly for a small to medium-sized company. A vendor with a cloud organization may support the critical computing capacity. A Mapreduce programming viewpoint isolates a large quantity of data [18]. From Mapreduce, an inquiry is conducted, and data are analysed to find key features to which the request is based; the findings are then limited to an application-noting dataset. The Mapreduce approach calls for the breakdown of large quantities of records. Each unique NAS contract is used for the mapping; simultaneous preparation is required for the mapping. Mapreduce's instant parallel requirements are costly and need the schedule for the cap. Expert cloud facilities can cope with needs.

6.2 Cloud Computing Services Models

The structures as an organization (PaaS), organization programming (SaaS), organization building (IaaS), and equipment (HaaS) are coupled with the models of specific organizations for transmitted transmission. Cloud-submitting game plans can deliver advantages that associations cannot monitor. Cloud transmission plans can also be used as a particular undertaking by affiliations before they grasp another application or significant progress. For PaaS cloud associations, the stage as a service requires appropriate registration to take steps towards customized applications and methods. PaaS intervention courses combine the mechanical assemblies' implementation strategy and lead-way, design processing, forming, consolidating, delivery and promotion, state administration, and associated improvements. Associations are using PaaS to create cost risk reserves by systematizing and using the cloud-based phase across different applications. The primary purpose when using PaaS includes reducing threats through pretested developments, advancing standard organizations, increasing the security of programming, and reducing mastery needs required for new structures to move forward. About extensive data, PaaS is giving associations a stage for making, utilizing, and examining large unstructured data initiatives rapidly and most of them in a safe and protected space, with anticipated custom applications.

The company does not pay hardware, only for the time and number of critical customer's information sharing limit. SaaS's key preference is to allow partnerships to interpret the threats found by the acquisition of programming while turning IT from reactive to constructive. The main fields for SaaS are smoother programming partnerships, updated changes, and repair organizations, market comparability, less mutual demands, and complete accountability. Programming as a service provides associations with programming responses to data examination, which separate extensive data. SaaS and PaaS optimize themselves in

this situation by not supplying SaaS with an altered intervention. At the same time, PaaS enables the group to create a solution specially customized to the organization's needs. For every reason behind the use of rigging to identify errors, including limits, equipment, servers, and organizational framework, the IaaS show will be used by a customer company. A distributed figuring model, with a desire for 25 percent of the efforts expected to achieve a community for IaaS, is an organizational system. Organizations that are open to IaaS organizations suggest that they assist in disaster relief, become a corporation, invest as a business, expand the server as an entity, and provide a global operating area and cloud effect program that offers the peak stacking capacity for variable strategies [19][20][22]. IaaS consolidates enhanced cash-related versatility, company preference, market resilience, strategic cleverness, and improved protection. Though not used as thoroughly as PaaS, SaaS, or IAS beforehand, HaaS has a cloud benefit in the timeshare model on mini-computers and built-in servers since the 1970s.

6.3 Record Encryption

Encryption ensures that consumer statistics are confidential and private while providing touching facts. Encryption prevents data if unauthorized users or managers take advantage of in-training exposure and automatically check out the documentation and eventually make unreadable compromised files or copied disk pictures. Data layer encryption provides constant protection on various platforms irrespective of the type of OS/platform. Encryption fulfils our huge mathematical protection criteria. For full Linux systems, Open Source solutions are available, but commercial software often has external key management and complete assistance. This provides an inexpensive avenue for specific data and health risks to be discussed.

6.4 Imposed Access Control

Authorization is the entry mechanism that determines the rights of an entity or a record security gadget for management. The encryption of file layers is not always useful if an invader can get encryption key entry. Many noteworthy facts encourage managers to keep their keys on nearby disks because it is comfortable and convenient. Still, it is also unsafe as keys can be gained via the policy administrator or an attacker. Using a key manager, keys and accreditations are preferred, and different keys are manipulated by each institution, software, and user.

6.5 Logging Issues

We need a list of activities to expose threats, evaluate accidents, or investigate insane behaviour. Whereas fewer flexible frameworks for the information and processing of periodic evidence, essential details are a perfect fit. Many web organizations, particularly for handling log records, begin with great information. It offers us a dimension of whether someone is lacking or whether anyone feels that something has become a hack. Therefore, it is necessary to test the whole system regularly to satisfy safety specifications. But secure operations will take time. The reality that much of the data may not be important to the task at hand every day is challenging to locate appropriate and accurate facts [21]. A business of tremendous data creates a difference every day between the truth and the figures of the

analyst. Significant statistics from Twitter can not necessarily representational evidence, even though all the details are loaded.

Moreover, an extensive collection of data will not require reliable figures every day. In certain circumstances, the higher the truth, the better the correct classifications. Big data units make unusual, critical incidents, more top-level analysis feasible. Large amounts of information can lead only to styles or connections without using the details of broader dynamics. In-house legal conclusions cannot be universal on a one-of-a-kind basis on an everyday basis by unconsulting samples. By random sampling, partial and non-represented examples are avoided. Statistics are not constantly additive, and the results of the subsets' assessment cannot be drawn every day. Units require overall scalability and performance in the processing of essential records. To make smaller record sets for analysis, data is usually filtered. The use of information requires identifying relevant and meaningful data, statistical value information, and the context and question requested to be understood.

The challenging statistical circumstances involve one-of-a-kind details kinds' common in regular, semi-structured, and unstructured data. A random fact represents an actual day-to-day record, and its yards in herbal mother tongue are conveyed and do not use a specified shape or area. Unstructured statistics created by humans are full of complexities, variances, and two definitions. Caution is needed when the material of random facts created by human beings is deciphered daily. Logical contradictions plague the assessment. By joining the glossary of business sentences, Hierarchy, and Taxonomies to corporate ideas, metadata can improve consistency.

7. Conclusion

Circumstances and needs described in three sizes of information: quantity, range, and speed. Creating an effective solution to broad and dynamic data is an endeavour that organizations in this sector continually study and endorse for successful management strategies. The disadvantage of open source is that it does not provide support and guidance like the paid programming. In this regard, all it is essential to maintain an immense structure for certainties and to work with practical needs is an out-of-entry design department in the most important occurrences. Equipment must be as fast as contemporary creativity may deliver. The software layout helps to account for a loss of feeling pace by planning and submitting orders so that processes can be optimized in general as an approach to first harvesting. Significant realities for appropriate use, which requires human assessment capacities, could be divided to manage measurements. Computer software should do what is tailored to; no hidden places can be found, so no different forms of ideas can be gained or adjusted unless they're updated. Human ability is thus necessary to type inevitabilities with a hardware company, which speeds up the procedure. The handiest blast as the findings will be seen would then minimize the recipient's optimal ability to take steps or prepare accordingly by evaluating the impacts, with a final aim to determine existing conditions or conjectures.

8. References

1. Paul Zikopoulos, Chris Eaton, and IBM. 2011. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data (1st. ed.). McGraw-Hill Osborne Media.

2. Alaaeddine Yousfi, Kimon Batoulis, and Mathias Weske. 2019. Achieving Business Process Improvement via Ubiquitous Decision-Aware Business Processes. *ACM Trans. Internet Technol.* 19, 1, Article 14 (March 2019), 19 pages.
3. Adesola, S. and Baines, T. (2005), "Developing and evaluating a methodology for business process improvement", *Business Process Management Journal*, Vol. 11 No. 1, pp. 37-46.
4. Arlbjørn, J. S. "Business Process Optimization/Jan Stentoft Arlbjørn, Anders Haug." Aarhus: Academica.–2010.–224 p (2010).
5. Amir Gandomi, Murtaza Haider, Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, Volume 35, Issue 2, 2015, Pages 137-144, ISSN 0268-4012.
6. Sangameswar, M.V., Nagabhushana Rao, M. & Satyanarayana, S. An algorithm for identification of natural disaster affected area. *J Big Data* 4, 39 (2017). <https://doi.org/10.1186/s40537-017-0096-1>
7. LaValle, Steve, et al. "Big data, analytics and the path from insights to value." *MIT sloan management review* 52.2 (2011): 21-32.
8. B. Renuka Devi, Dr. K. Nageswara Rao, Dr. S. Pallam Setty, Dr. M. Nagabhushana Rao. "Disaster Prediction System Using IBM SPSS Data Mining Tool". *International Journal of Engineering Trends and Technology (IJETT)*. V4(8):3352-3357 Jul 2013. ISSN:2231-5381.
9. Hu, Han, et al. "Toward scalable systems for big data analytics: A technology tutorial." *IEEE access* 2 (2014): 652-687.
10. Rao, NK Kameswara, Dr. GP Saradhi Varma, and Dr. M. Nagabhushana Rao. "Spatial Mining System for Disaster Management." *International Journal Of Innovative Technology And Research*, Volume 1: 033-036.
11. Najafabadi, Maryam M., et al. "Deep learning applications and challenges in big data analytics." *Journal of Big Data* 2.1 (2015): 1.
12. Kaleemullah, T., et al. "Sensitive ion chromatographic determination of citrate and formate in pharmaceuticals." *Rasayan J Chem* 4 (2011): 844-852.
13. Kaleemullah, T., et al. "Sensitive ion chromatographic determination of citrate and formate in pharmaceuticals." *Rasayan J Chem* 4 (2011): 844-852.
14. Praveen, S. P., Nguyen, H. H. C., Swapna, D., Rao, K. K. and Kumar, D. L. S. (2020). The Efficient way to Detect and Stall Fake Articles in Public Media using the Blockchain Technique: Proof of Trustworthiness. *International Journal on Emerging Technologies*, 11(3): 158–163.
15. Praveen, S. P., Rao, K. T., & Janakiramaiah, B. (2018). Effective allocation of resources and task scheduling in cloud environment using social group optimization. *Arabian Journal for Science and Engineering*, 43(8), 4265-4272.
16. R. Arun Prakash, T. Jayasankar, K. Vinoth Kumar, "Biometric Encoding and Biometric Authentication (BEBA) Protocol for Secure Cloud in M-Commerce Environment", *Appl. Math. Inf. Sci.* Vol.12, No.1, Jan 2018, pp.255–263. DOI: <http://dx.doi.org/10.18576/amis/12012>.
17. Swapna, D., & Praveen, S. P. (2019, October). An Exploration of Distributed Access Control Mechanism Using Blockchain. In *Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics (Vol. 2, p. 13)*. Springer Nature.
18. Praveen, S. P., & Rao, K. T. (2016). An Algorithm for Rank Computing Resource Provisioning in Cloud Computing. *International Journal of Computer Science and Information Security (IJCSIS)*, 14(9).
19. S. Pramela Devi, V. Eswaramoorthy, K. Vinoth Kumar and T. Jayasankar (2020), Likelihood based Node Fitness Evaluation Method for Data Authentication in MANET, *International Journal of Advanced Science and Technology*, vol. 29, no. 3, pp. 5835 – 5842.
20. Praveen, S. P., & Rao, K. T. (2019). An Effective Multi-faceted Cost Model for Auto-scaling of Servers in Cloud. In *Smart Intelligent Computing and Applications* (pp. 591–601). Springer, Singapore.
21. Praveen, S. P., Tulasi, U., & Teja, K. A. K. (2014). A cost efficient resource provisioning approach using virtual machine placement. *Int. J. Comput. Sci. Inf. Technol.*, 5(2), 2365-2368.
22. M. Anuradha, T. Jayasankar, Prakash N. B., Mohamed Yacin Sikkandar, G. R. Hemalakshmi, C. Bharatiraja & A. Sagai Francis Britto (2021), IoT enabled Cancer Prediction System to Enhance the Authentication and Security using Cloud Computing, Microprocessor and Microsystems, Vol 80, February, 103301 <https://doi.org/10.1016/j.micpro.2020.103301>.