Big Data & Cloud computing Based Analytical Tool for Digital India

Prof Ranjeet Ramesh Pawar[1], Prof Mithun Vishnu Mhatre[2]

HOD, Information Technology Department, B.V.I.T Navi Mumbai[1]

HOD, Computer Technology Department, B.V.I.T Navi Mumbai[2]

Abstract: Big data is an emerging and popular trend in information technology. Operation on data and analysis is usually a remarkable thing which was accustomed keep records and data for future purpose. the majority sectors required data collection and that they perform different operations thereon so this technology accepted by all. As data collected from any wired or wireless media so to stay safe and accurate is that the main role in concern with security. Data stored within the cloud always give flexibility to access from anywhere on any platform. Real time operation are often performed on data through cloud to create digital applications. Big data and cloud computing move to serve input furthermore as output for several applications like education, medical, government, etc. As we see dream of digital India so these two things give more contribution towards it. Maintaining security of information over cloud is required at a high level through security techniques. Maintaining integrity of information could be a key challenge to save lots of sensitive data on cloud. This paper covers all points regarding concept of massive data, analysis, operations, tools for data collections and applications of huge data. It also covers the importance of huge data concept for digital India. This paper also deals with cloud security and techniques accustomed secure data on the cloud.

Keywords: Big Data, Analytics, Cloud, Security, Digital India**.**

I. INTRODUCTION

1. CONCEPT OF BIG DATA

Beginning of an enormous data era started when data is generated at an implausible speed from everywhere. Computing has become global, number of devices like cell phones, smart phones, laptops, personal sensors; social sites are creating countless new digital oceans of knowledge. Big data can be a term that refers to datasets that are so large or complex that traditional processing application software is inadequate to handle them. The three V volume, speed and variability of the large data made by Doug Laney in 2001 to handle the info management challenge was able to define big data for some years. It basically interprets big data like plenty of knowledge in scattered form and desires of knowledge that's in a very scattered form and desires to be processed quickly for correct interpretation. Big Data mainly involve six aspects as per figure- 1 as per the mentioned definition.
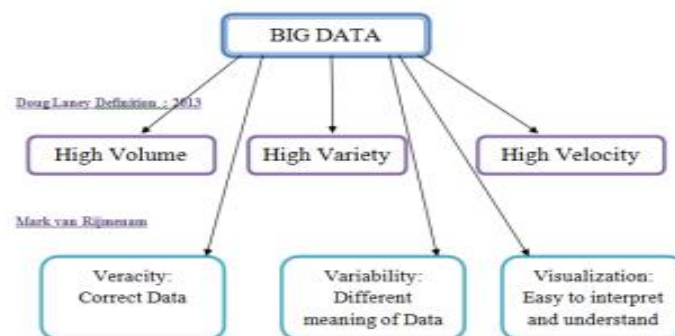

Figure1. Characteristics of Big Data

- **Volume**: Volume defines the number of massive Data. the dimensions of this data ranges from terabytes and petabytes, to even Exabyte's.

- **Variety**: Variety define data varieties of Big Data, which incorporates structured and unstructured data like text, audio, video, sensor data, posts, log files and plenty of more.
- **Velocity**: because the generation of knowledge is rapid, the method of acquiring, processing and analyzing it requires fast mechanisms. Speed emphasizes real-time performance of massive business needs data.
- **Veracity**: Refers to the necessity of correct kind of data because it is relied upon for all further analysis.
- **Variability**: Data is within the same form but having different semantics.
- **Visualization**: Data should be easy to process and interpret to derive intelligence out of it.

2. ANALYSIS OF BIG DATA

The analysis of massive Data involves multiple distinct phases as shown within the figure2:
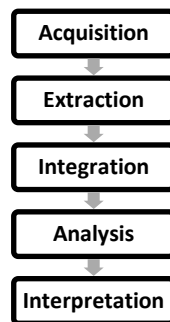

Figure2. Big Data analysis phases

- **Acquisition**: big Data doesn't arise out of a vacuum: it's recorded from some data generating from different sources.
- **Extraction**: the data collected won't be during a format ready for analysis. the method of extracting information that pulls the desired information from the first sources and presents it in a very systematic manner merit analysis.
- **Integration**: Data integration involves combining data from several disparate sources, which are stored using various technologies and supply a unified view of the info.
- **Analysis**: Big Data is additionally enabling the following generation of interactive data analysis with real-time answers. within the future, Big Data-directed queries are generated automatically to form content.
- **Interpretation:** Having the flexibility to investigate Big Data is of limited value if users cannot understand the analysis. Finally, the person making the choice, given the results of the analysis, must interpret these results.

II. CHALLENGES & HANDLING BIG DATA

- In engineering, Big Data could be a sensitive issue that needs attention. There are only two or three main issues appear capable of constructing or breaking the promise of huge Data, and these are related to: solution approach, personal privacy and intellectual priority (IP). Challenges of massive Data related concern is data securities are:
- **Heterogeneity:** Machine analysis algorithms expect homogenous data, and can't understand nuance. As a result, data should be carefully sorted as a primary step (or before) data analysis.
- **Scale:** As we've examined above that volume or scale is that the second major challenge of massive Data. Managing large and fast-growing data has been a challenge for many years. Data volume is scaling faster those computer resources.
- **Timeliness**- it's directly related with size, larger the scale of knowledge longer is required to process and analyze data.

- **Personal Privacy:** Maintaining secure information which obtained through various resources or which are stored on various platform like cloud.

Handling of massive Data is another major concern. There are some emerging technologies that are helping users deal with and handle Big Data in an exceedingly cost-effective manner. Big data handling is through with relevancy following aspects:

- **Processing Big data**: MapReduce, Hadoop is an integrated framework for processing and storing Big data
- **Analysis and querying of data**: WibiData, PLATFORA, PIG
- **Business Intelligence**: Hive
- **Storage**: Cloud storage, Column-oriented databases, schema-less databases
- **Machine Learning**: Apache Mahout, SkyTree.

III.APPLICATIONS OF BIG DATA

Companies have worked on data analysis and translation to open up new opportunities in their own way. Big data represents important details and the way they'll be analyzed. This provides a chance at every stage of data discovery in big data. Big data provides a chance in many areas as mentioned below:

- Banking and security
- Communication Media and Services
- Government
- Insurance
- Manufacturing and natural resources
- Transportation
- Various Traders such as: Retail, wholesalers

Big data in Education: There's a robust trend for enormous Web deployment of educational activities, and this can generate an increasingly great amount of detailed data about students' performance. So as to attain accreditation this data will be useful to figure with open data. Big Data can really improve the education. It can afford to shape a contemporary and dynamic education system, which each and every individual student can have the utmost have the benefit of that. Additionally, teachers have important tools, if they didn't have them before, which will make their decisions clearer and that they can choose different learning styles. Hence the massive Data are literally involved to vary the way of industries including the education. Within the new era of knowledge the normal difficulties are going to be now not exists, keeping the nice methods. The education system are going to be enriched with new learning ways, making more efficient and targeted. However, the path to this new age has only just begun, and there are several challenges ahead, such as a scarcity of qualified professionals in the science of large data and data analytics. Furthermore the teachers and academics must actually train and involved on them and at last the scholars must accept and use these new tools.

Big Data in Healthcare: It's widely believed that the utilization of knowledge technology can reduce the value of healthcare while improving its quality by making care more preventive and customized and supported on ongoing awareness (home-based).

IV.BIG DATA & CLOUD

Big data is to blame for storing and processing data, cloud provides a reliable, fault tolerant, available and scalable environment in order that big data systems can perform. Cloud computing provides a reliable, fault-tolerant, available and scalable environment to harbor astronomically immense data distributed management systems. Within the context of this paper we present an summary of both technologies and cases of success when integrating big data and cloud frameworks. Despite the fact that big data solves many of our existing problems, there are still certain gaps and challenges that cause worry and require development. Other concerns, including as security, privacy, scalability, data governance principles, data heterogeneity, and disaster recovery techniques, have yet to be resolved. Main issue is Privacy & Security.

Cloud computing and massive data security are this and important topics of research. This Problem becomes a difficulty to corporations when considering uploading data onto the cloud. Questions like who is that the real owner of the information, where the information, who has access to that is and what reasonably permissions they need are hard to explain. Corporations that are getting to do business with a cloud provider should bear in mind and ask the subsequent question:

    1. **Who is that the real owner of the information and who has access to it?**
    2. What about the data?

Through cloud a call cum feedback system has been developed to get report, precautions supported user health information at low cost. System is operated through mobile and supply good marketplace for health service provider. Privacy of user health information and therefore the properties of health service provider kept to be secure in cloud. New system required to keep up privacy with less overhead al client side for secure computation. Programs through which feedback or decision reply to user kept to be secured from inside and outdoors attacks. Decryption outsourcing and newly proposed secure key duplicate 2-way encryption are selected to cut back overhead at client side with securing privacy of knowledge. This paper demonstrates the effectiveness of system in terms of security and computational performance.
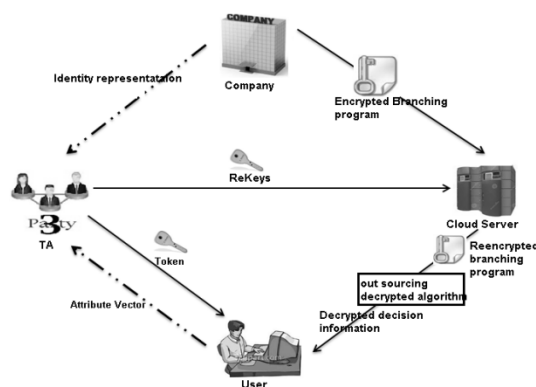
V. PROPOSED SYSTEM:

Figure3. Proposed system for data security

A cloud predicated health monitoring system includes four components or parties i.e. Utiliser at the client-side, Cloud server, Health accommodation provider, Third-party auditor. Figure 3 shows the architecture of the proposed system without utilizing incipiently proposed secure key duplicate 2-way encryption which not gives performance analysis in terms of security and efficiency. An advance system which uses secure key duplicate 2-way encryption for security and efficiency during which 2 ways encryption is completed. It also reduces computational workload at user side and move to cloud. During initialization third party auditor initializes and run setup phase and it generates system required parameter. After generation of system parameter it'll publish it to user, cloud server and health service provider. After initialization health service provider will store there medical applications programme within the variety of branching tree in cloud. This branching tree is encrypted and generated cipher text is stored in cloud. to spot accommodation provider, each accommodation provider gets one index and together with that index encrypted branching tree program stored in cloud. When particular utiliser wants decision or feedback from accommodation provider it commences token generation operation in sodality with third party auditor. Client sends index value of health service provider together with its input value vector which consists of user health information. User input query passed within the sort of vector with information components. After getting input query from client, Third party auditor generates one token and sends to client. During this process third party won't get any user identifiable information. The client transmits a token to the cloud server, which is needed to make a decision. Third party auditor will validate tokens and cloud send this token to health accommodation provider then accommodation provider will engender feedback or decision fortified decision tree structure and it absolutely was passes to cloud in cipher text format. The client will get partially decrypted cypher text from the cloud. Because this is only a partial communication, the cloud will not get any useful decision or response.

VI. DIGITAL INDIA/ SMART INDIA

In India, the **"Digital India"** initiative was launched with the goal of transforming the country into a digitally empowered society and knowledge economy. Its vision is broadly centered around three key areas: i) Infrastructure as a utility for each person, ii) Governance and on-demand accommodations, and iii) Digital denizen potentiating. Consequently availability of assorted Data resources has increased. These data resources can provide a chance to the business enterprises to explore the business trends and switch investments into huge profitable returns. so as to provide better e-Governance service to citizen it's important to use big data technology effectively. Use of this technology are often done before launching any service or after launching any service. Big thing was required for better India, Digital India or Smart India which is Big Data.

VII.   CONCLUSION

With data increasing on a daily base, big data systems and specifically, analytic tools, became a significant force of innovation that has some way to store, process and acquire information over petabyte datasets. Big data solutions are heavily reliant on cloud settings, which provide fault-tolerant, scalable, and available settings for large systems. Big Data can really improve the education. It can afford to shape a contemporary and dynamic education system, which each and every individual student can have the most get pleasure from that. Big data is employed in all the sectors and now it becomes basic need of any organization or industry. Proposed encryption technique will secure data on cloud and increase privacy of data. Big data also strengthen government programs for better India. Various future and existing challenge must be solved for better use of technology.

REFERENCES

[1] González-Martínez, J. a. et al., 2015. cloud computing and education: A state-of-the-art survey. Computers & Education, 80, pp.132–151.

[2] Hashem, I.A.T. et al., 2014. The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, pp.98–115.

[3] Smart Health and Wellbeing. Computing Community Consortium. Spring 2011.

[4] Jean Yan, April 9, 2013 ―Big Data, Bigger Opportunities Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems.

[5] A Review of Big Data Management, Benefits and Challenges 1 Oguntimilehin A., 2 Ademola E.O. 1,2Department of Computer Science, AfeBabalola University, Ado-Ekiti, Nigeria