Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 6 – No.1, October 2016 – www.caeaccess.org

Security and Privacy Preservation on Cloud-based Big Data Analysis (CBDA): A Review

Hemanth Kumar N. P.
Assistant Professor
Department of CS&E
AIET, Moodbidri, India

Prabhudeva S.
HOD,
Dept. of ISE, JNNCE, Shivamogga, India

ABSTRACT

The term Big Data is nothing but large voluminous data, more complex data and relationship analysis of these data sets. The today's organizations, business units, government sectors, etc. are adopting the big data technique to store the enormous data generated by them. With all the significances of the big data as per economic and social benefits, lacks with many issues related to the security and privacy of the data. The modern ethnology like cloud computing will offer a scalable service for the big data with optimized cost. But the concern of privacy and security is still unsolved. This paper reflects the survey over the cloud-based big data security and privacy preservation. The survey discusses the recent work carried for privacy preservation and also existing research gap. The survey states a significant section for the future research line-up.

Keywords

Big data, Cloud Computing, Privacy and security issues

1. INTRODUCTION

The term big data is nothing but large voluminous data, more complex data and relationship analysis of these data sets. The main advantage of big data is that it performs the better analysis of huge data than conventional analysis methods. Due to this reason the big data has gained very much interest in the present generation, which has advancement in the data collection, data storage and performs the data interpretation. From last few decades, the use of digital media is been increased in many areas which generating the tremendous amount of data, for example, hospital data, bank data, social networking data, etc. The data storage cost is decreasing day by day by which we can store the entire data rather than discarding it. In addition to this, many of the data analyzing techniques are developed and have not succeeded for efficient data analysis. The cloud computing is the recent technique which offers many significant advancements in the research work of the big data analysis. In this way, cloud arranges enormous data that contain sensitive information and are required to send particular measures and various leveled shields to keep up a key separation from data confirmation breakdowns that may achieve tremendous and extravagant damages. Critical information as to disseminate figuring incorporates data from a broad assortment of different locales and requests. Over the time, affiliations have assembled noteworthy information about the general population in our social requests that contain tricky information, e.g. therapeutic data. Researchers need to get to and analyze such data using huge data propels as a piece of circulated processing, while affiliations are required to actualize data security consistence. There has been huge progression on privacy protection for fragile data in both industry and the informed group, e.g., plans that make traditions and mechanical assemblies for anonymization or encryption of data for security purposes. This portion sorts business identified with this area according to assorted security protection necessities. Regardless, these courses of action have not yet been for the most part gotten by cloud organization suppliers or affiliations. Like this, with the extension of these new cloud headways recently, security and data protection requirements have been creating to guarantee individuals against surveillance and database exposure.

Cloud computing has raised several security threats such as, malicious insiders, data loss, data breaches, and denial of service that have been extensively studied. These threats mainly originate from issues such as multi-tenancy, loss of control over data and trust. This means that there are important concerns about security and privacy need to be focused on cloud computing by all parties involved in the cloud computing arena.

This paper discusses some significant aspects of the big data over cloud computing with the privacy and security issues. The most significant and latest existing work towards the privacy and security solution in cloud-based big data analysis (CBDA) is discussed and finalized with the future level of the work required in the CBDA privacy and security solution. The section wise representation of this survey paper is provided as Section 2 followed with the big data concepts and security and privacy issues; Section 3 represents the cloud computing concepts and privacy & security issues in it, methods used for cloud-based big data analysis. Section 4 talks about the existing research work are CBDA and privacy & security preservation. Section 5 figure out the research gap in existing research gap. Section 6 suggests the future research lineup and the conclusion are given in section 7.

2. BIG DATA

The term big data is nothing but large voluminous data, more complex data and relationship analysis of these data sets. The main advantage of big data is that it performs the better analysis of huge data than conventional analysis methods. Due to this reason the big data has gained very much interest in the present generation, which has advancement in the data collection, data storage and performs the data interpretation. From last few decades, the use of digital media is being increased in many areas which generating the tremendous amount of data, for example, hospital data, bank data, social networking data, etc. The data storage cost is decreasing day by day by which we can store the entire data rather than discarding it. In addition to this, many of the data analyzing techniques are developed and have not succeeded in efficient data analysis.

The big data in the real world is like the collection of huge resources which can be used regularly. The big data provides the vast application advantages, but the conventional data analysis methods fail to provide the proper privacy mechanism. The privacy concern of the big data includes the private data disclosure to the world.

2.1 Big Data Characteristics

The big data helps in storing, analyzing and processing of huge volume of data. Some of the big data characteristics are addressed below:

2.1.1 Characteristic 1-Volume

The volume of big data means the collection of the large volume of data from various data sources and continues data expansion. The large Volume data collection helps in providing the better-hidden pattern via data analysis mechanism.

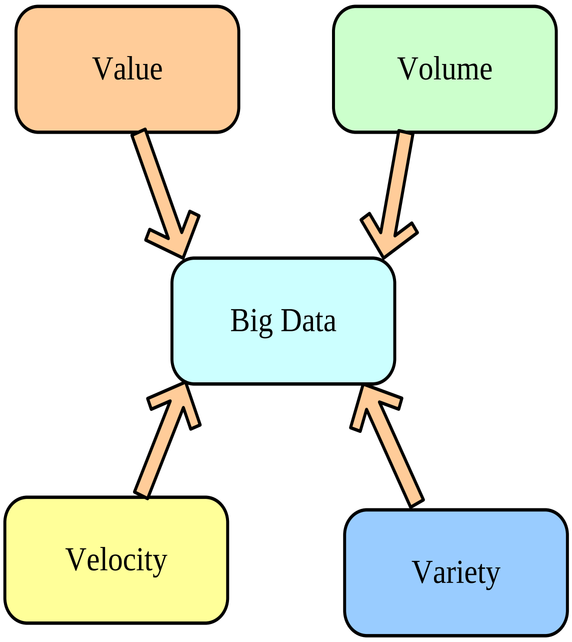

Figure.1 Four V's of Big Data

2.1.2 Characteristic 2-Variety

The variety in big data collection is the collection of various types of data from different sources like smart phones, sensor nodes, social networks, etc. All these data sources are of digital media data comprised of image, video, audio or data logs. The internet based structured and unstructured data is more among the big data variety.

2.1.3 Characteristic 3-Velocity

The velocity in the big data is the speed of the data file transfer rate. In real time, the data transfer speed varies with the variation in data collection from different sources.

2.1.4 Characteristic 4-Value

The value in big data referred as the finding of various hidden values from huge data sets of different types and faster generation of data.

2.2 Big Data Types

The Big data can be categorized into various types that help in better understanding of the BD characteristics. This categorization of BD provides the better understanding of large data stored in the cloud system. The types are: data sources type big data (BD), data processing type, format content type, data staging type, and data stores type and are explained as below.

Type 1- Data Sources: The data sources type of big data includes social media data, machine data, sensing data, transactions data and IoT data.

a. ***Social media data:*** This is data generated by many social blogs, microblogs, Facebook, Twitter through the virtual communication.

b. ***Machine data:*** This is data generated by some hardware's or software's based on hospital machines, computers without humans help.

c. ***Sensing data:*** This is data gathered by sensors and converts them as signals.

d. ***Transactional data:*** The data's like bank transactions, work records, and other time dimension based events data are considered as transactional data.

e. ***Internet of things (IoT) data:*** The data generated by the interface of Smartphone's, tablets and digital cameras with the internet are considered as IOT data.

Type 2- Data processing data: This type of data includes real-time and batch data.

a. ***Batch data:*** Many organizations in recent days are adopted the Map Reduce systems for running the batch jobs, such systems offers the large data scaling generated from many nodes.

b. ***Real-time data:*** The data generated with the real-time computation, processing, etc.

Type 3- content format data: The content format data types are divided into structured, semi structured and unstructured data.

a. ***Structured type data:*** These kinds of data are handled with SQL. The data is easy to apply as input, for storage, query and analysis purpose. Numbers, dates, and words are the structured data.

b. ***Semi-structure type data:*** These are may be of structured and are not organized as rational databases or tables.

c. ***Unstructured type data:*** This data includes location information, texts, videos and social data.

Type 4- Data staging: The data staging includes cleaning data; transform data, and normalization data.

a. ***Cleaning data:*** This is unreasonable, incomplete data identification or cleaning.

b. ***Transform data:*** The process data were transforming as the suitable form of data for analysis.

c. ***Normalization data*****:** This is database structuring method to reduce the redundancy.

Type 5- Data stores: This is divided into document oriented type of data, Column-oriented data, graph database, key-value type, and all are used to store the data.

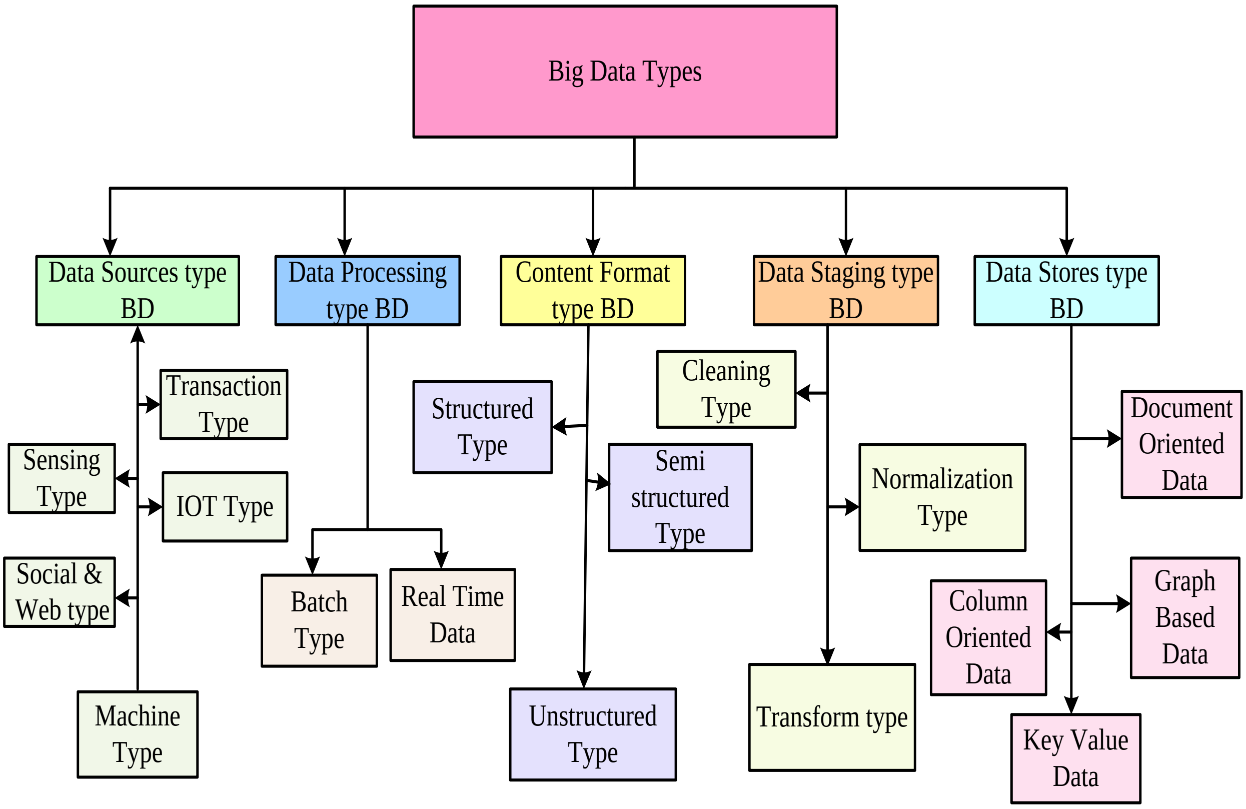

Figure.2 Big Data Types

2.3 Cloud Computing

Today's world of IT industry is more effective with the cloud computing (CC) technology. The premise of CC includes infrastructure as a service (IaaS), software's, and hardware's and these can be stored in the remote data center through the internet. The CC provides architecture for computing the large scale data issues and sorts the issues related to the data storage in IT industries. There are many factors has made IT industries to face towards cloud computing for storage, analysis, and processing purpose. The today's scientific and other research centers are adopting CC to store their experiments over the cloud. But the issue is that these experimental data is increasing day by day. Also, the cloud service providers are integrating the many different frameworks to make the users feel free in accessing cloud resources and deploying their data/ programs in the cloud server. The CC exhibit many service unit and are explained as below.

2.4 Service model of cloud computing

The service model is divided into three service units as infrastructure as a service, platform as a service and software as a service.

Unit 1: Infrastructure as a service (IaaS): The unit provides the significant service to connect the virtual network in short time, and the consumer can pay the money for the resources which he has used. The unit has significant parts like servers, peripheral devices, and storage devices. This unit can be interconnected with managed service. This unit offers consumers to manage the infrastructure by web based GUI (Graphical User Interface), and it serves as console management for IT operation.

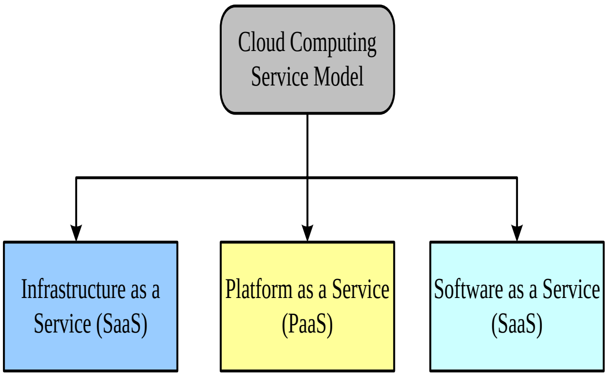
Cloud Computing Service Model
Infrastructure as a Service (SaaS)
Platform as a Service (PaaS)
Software as a Service (SaaS)

Figure.3 Cloud Computing Service Model

Unit 2: Platform as a service (IaaS): This is used by the developer to instruct or write the commands which execute on the cloud. This service is helpful for development and developers point of view. The service is highly scalable in nature. Example: Salesforce.com and Azure. The PaaS Offered the development platform for both in progress and completed while the SaaS offers service for the completed Cloud application. Paas model offers higher abstraction level. In this, the consumer can create software using libraries or tools of the provider and also the consumer control the settings of software configuration and deployment. The provider offers the servers, networks, storage, and other services for consumer's application.

Unit 3: Software as a service (SaaS): This licenses the software application of the provider to be used and purchase on demand. SaaS run on cloud along with many end users. E.g. Gmail- a popular SaaS product and it runs on browsers. I.SaaS applications similarly support what is traditionally known as application customization or like traditional enterprise software, a single customer can change the configuration options. Every customer will have own settings for the configuration options. The application can be customized to the degree it was designed for based on a set of predefined configuration options.

2.5 Cloud-based Big Data Analysis Architecture

The combined architecture of cloud system and big data system are shown in figure 3. Huge information sources from the cloud and Web are put away in a circulated shortcoming tolerant database and handled through a programming model for extensive data sets with a parallel dispersed calculation in a group. The primary motivation behind information representation is to see explanatory results displayed outwardly through various diagrams for basic leadership. Enormous information uses disseminated capacity innovation taking into account distributed computing as opposed to nearby stockpiling appended to a PC or electronic gadget. Big data assessment is driven by quickly developing cloud-based applications created utilizing virtualized advancements. In this manner, cloud computing not just gives computation and processing of enormous data additionally serve as a service model.

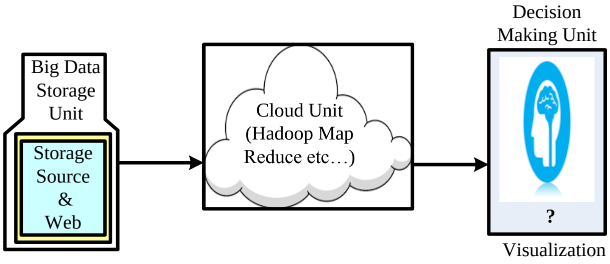

Figure.4 Architecture of Cloud-Based Big Data Analysis

3. SECURITY AND PRIVACY ISSUES IN CLOUD-BASED BIG DATA ANALYSIS AND PROTECTION TOOLS

3.1 Security and privacy issues

The security and privacy issues are given as below:

Architectural or Distributed Node type issue: The computation of any distributed node set is done, and data is processed with desired resources. The distributed node issues may occur anywhere in the cluster, and the identification of its computation is very difficult. Hence the security ensuring in this condition is quite a tough task.

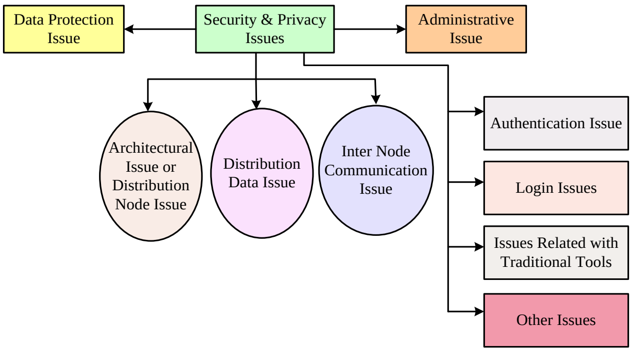
Data Protection Issue
Security & Privacy Issues
Administrative Issue
Architectural Issue or Distribution Node Issue
Distribution Data Issue
Inter Node Communication Issue
Authentication Issue
Login Issues
Issues Related with Traditional Tools
Other Issues

Figure.5 Security & Privacy Issues in CBDA

- ***Internodes' Communication issue:*** Many of Hadoop distributions transfer the data between the nodes by using RPC over the IP. This process may take place over the wired or wireless network, by which anyone can tap and break the communication of the system by modifying the internodes communication.
- ***Data protection issue:*** In many of the cloud environment the data is stored without encryption, which is done to enhance the efficiency. This will not predict or stop the hackers access the critical data.
- ***Administrative rights for nodes:*** The administrative of any node will have accessibility in accessing any data. This unconditional access will be very dangerous like any malicious by which the data can be manipulated or steeled.
- ***Nodes and applications authentication issue:*** The joining of clusters by the help of nodes may cause the increment in parallel operation and there is no authentication is provided the third part nodes may join at any time and steal the cluster data.
- ***Logging issue:*** The absence of logging feature in the cloud may lead the no record of activities of user data modification or delete. By this reason, any user or hacker can change/manipulate the data with ease.
- ***Issues of traditional security tools:*** The traditional tools are not scalable to the huge cloud system and hence these are not being efficient in the scalable cloud system.
- ***Issues related to the different cloud technologies:*** Use of may cloud technologies may lead in weaker cloud security.

3.2 Methods used for cloud big data security and privacy protection

The existing methods for cloud big data are described as below:

- ***File encryption method:*** The critical data in the cluster can be stolen by hackers and hence this data is needed to be encrypted properly. There exist various keys for encryptions and are used depending upon its applicability. The encrypted file cannot be hacked easily, and even the hacker hacks the file, but he will not be able to extract the meaningful data and misuse it. The data will be stored fully secure manner.
- ***Network encryption method:*** Every communication networks are needed to be encrypted as per the standards. The Remote Procedure Call (RPC) will help in protecting the network packets from the hacker.
- ***Logging method:*** In this, the authenticated user data can be stored which will help in auditing the malicious operations performed by the unauthorized used. This will help insecure logging process.
- ***Node and software format maintenance method:*** In the existing nodes of the software will be formatted continuously and accordingly the viruses are formatted. The keeping update of Hadoop software will offer extra security.
- ***Node authentication method:*** When the node is connected to the cluster it has to authenticate. If there is, any malicious node exists it will not access the cluster. The Kerberos is the most used node authentication protocol.
- ***Rigorous testing method:*** The developed Map Reduce job needs to be tested rigorously to bring the stability in it.

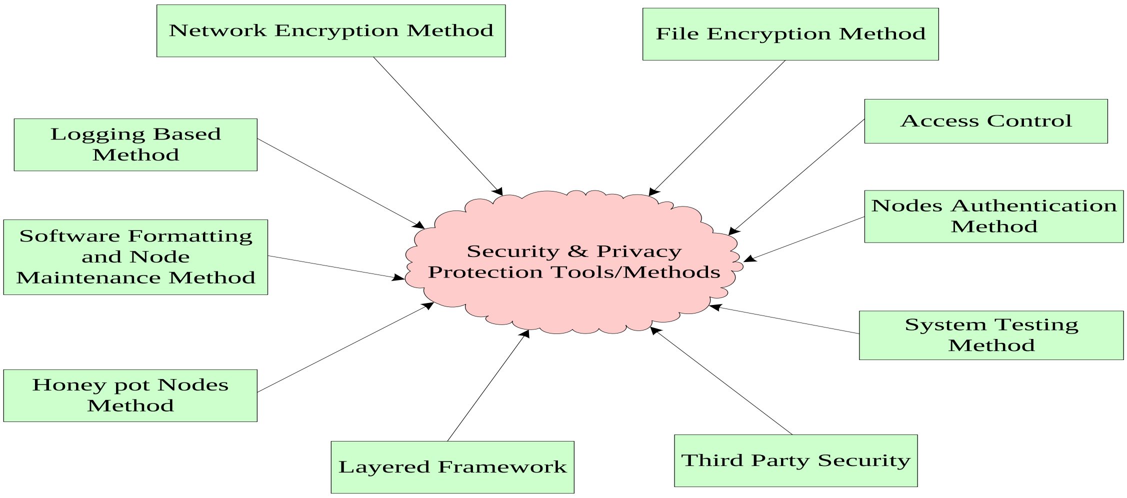
Network Encryption Method
File Encryption Method
Access Control
Logging Based Method
Nodes Authentication Method
Software Formatting and Node Maintenance Method
Security & Privacy Protection Tools/Methods
System Testing Method
Honey pot Nodes Method
Layered Framework
Third Party Security

Figure.6 Security & Privacy Protection Tools-Methods

- ***Honey pot method:*** In a cluster, there will be Honey pots which will trap the hacker and removes the hacker by performing the necessary process.
- ***Layered framework method:*** The following figure 7. Represents the layered framework method to solve the security issues in cloud computing. The framework has significant parts like user interface, secure cloud data manager, and storage manager, secure virtual monitor layer and cloud monitor layer.
- ***Third party control method:*** Cloud system will help in storing the data in a remote area, and the protection in this area is very difficult. The third party will take the overall responsibility for the secure transaction from both the owner side and consumer side.

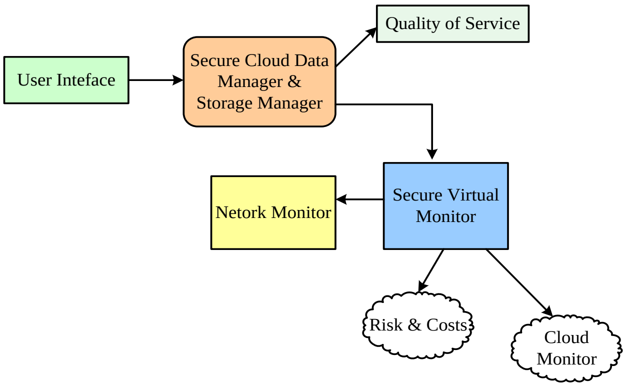

Figure.7 Layered Framework Method

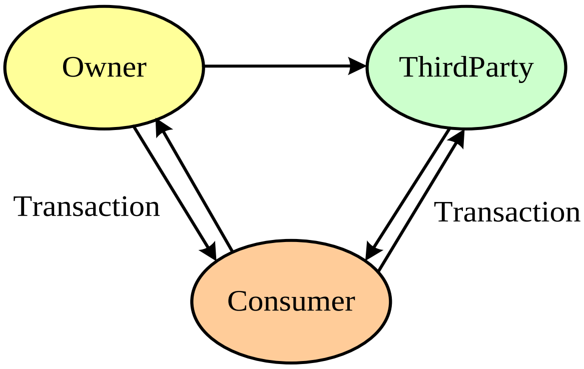

Figure.8 Third Party Secure Method

- ***Access Control method:*** the proper access control mechanism will provide the secure data protection.

4. EXISTING WORKS OVER PRIVACY AND SECURITY ISSUES OF BIG DATA ANALYSIS BASED ON CLOUD

The section literally explains the existing research work and is given below.

Mayilvaganan and Sabitha [1] have introduced the CBDA in smart grids to minimize the power usage. The author has driven energy from renewable sources and significantly explained the future scope with the power demands.

Baciu and Zhang [2] presented the cloud-based data analysis for large streaming data known as Claudette. For the issues related to the parallel computation hierarchical cloud architecture is described. At next author illustrated the stream data mechanism to minimize the stream. In the end, the visual cognition mechanism to detect the data set salient regions. Tamura et al. [3] have illustrated the 3D stochastic differential equation based reliability analysis for CBDA. The analysis method is implemented to find out the network factor, big data factor, and fault. The study assumed the software detection rate Vs time. Guo et al. [4] have described the placement algorithms for BD in the private cloud. The method significantly focused on the I/O data balancing and outcomes with better potential results. Pant et al. [5] presented the three step cloud data security by encryption methods. The method helps in secure data storing mechanism of steganography and cryptography. Luo et al. [6] give the aspects of distributed big data storage over the cloud. The method implemented the Boafft system prototype over Hadoop distributed file system. Lemoudden and El Ouahidi [7] given the big data method for Logs data generated by cloud. The log files are mainly for the troubleshooting, debugging, security and compliance, etc. management. Adnan et al. [8] presented the concept of issues optimization for big data based on Hadoop-based cloud architecture. The method has got improved performance. Banditwattanawong and Uthayopas [9] presented the sharing of big data with i-cloud in point of efficiency and economic point of view. The trace-driven simulation results cost optimized and efficient of data sharing. Spicuglia et al. [10] have given the novel solution for big data cloud data allocation to enhance the capacity data storage. The method provides the robust way to store the huge data over the cloud.

Sharif et al.[11] Given the survey over the existing security threats and preventive measures of cloud computing, Hadoop, and big data. The survey has performed the security analysis over EnCoRe system and suggested the preventive measures for it.

Essa et al. [12] presented the Hadoop-based parallel and distributed big data analysis. The study has focused on the main drawbacks of Hadoop system and out formed a new big data analysis model known as Map Reduce Agent Mobility (MRAM) over the mobile and Map reduce protocol developed in Java language. The method outcomes with the efficient results and solves the issues related in Hadoop.

Cau et al. [13] described the detection method of service Denial based on Entropy for the cloud data centers. Authors have analyzed the attack of malicious over the virtual machines and developed the entropy based method to monitor the threats on the virtual machine and outcome with the significant data center results.

Zheng et al. [14] given the internet of vehicles based method of big data analysis over the cloud computing. The authors have surveyed the issues in big data incorporated with the IOV and improve the IOV capability.

Gohil et al. [15] proposed the performance analysis factors of the Map Reduce applications in cloud-based Hadoop for big data. The author experimented on cloud-based Word count, Terasoft, Grep, pi and Hadoop applications and analyzed the performances like the number of nodes and execution time. The conclusion of the authors work states that the work decreased the execution time and enhanced the performance.

Table 1 Summary of existing research literature

Author	***Contribution***	***Type***	***method***	***Remarks***
Mayilvaganan and Sabitha [1]	Research proposal for cloud-based big data analysis	Conferenc e	Cloud architecture	Proposed the architecture use in smart grids
Baciu and Zhang [2]	Cloud-based data analysis for streaming data	IEEE transactio n	Cloud-based method	• For the issues related to the parallel computation hierarchical cloud architecture is described. • At next author illustrated the stream data mechanism to minimize the stream. • In the end, the visual cognition mechanism to detect the data set salient regions.
Tamura et al. [3]	3D stochastic differential equation based reliability analysis for CBDA	IEEE transactio n	Cloud-based big data analysis	The analysis method is implemented to find out the network factor, big data factor, and fault. The study assumed the software detection rate Vs time.
Guo et al. [4]	The placement algorithms for BD in private cloud for I/O load balance	IEEE transactio n	Big data analysis based on private cloud	The method significantly focused on the I/O data balancing and outcomes with better potential results.
Pant et al. [5]	The three step cloud data security by encryption methods	IEEE transactio n	Encryption and steganography method	The method helps in secure data storing mechanism of steganography and cryptography
Luo et al. [6]	The aspects of distributed big data storage over the cloud	IEEE transactio n	Boafft system prototype	secure data storing mechanism of steganography and cryptography
Lemoudden and El Ouahidi [7]	Big data method for Logs data generated by cloud	IEEE transactio n	Survey	Survey over data management
Adnan et al. [8]	The concept of issues optimization for big data based on Hadoop-based cloud architecture	IEEE transactio	Hadoop-based cloud computing for big data issues	The method has got improved performance.

		n		
Banditwattana wong and Uthayopas [9]	the sharing of big data with i-cloud in point of efficiency and economic point of view	IEEE transactio n	i-cloud	The trace-driven simulation results cost optimized and efficient of data sharing
Spicuglia et al. [10]	The novel solution for big data cloud data allocation to enhance the capacity data storage	IEEE transactio n	Optica method	The method provides the robust way to store the huge data over the cloud.
Sharif et al.[11]	survey over the existing security threats and preventive measures of cloud computing, Hadoop, and big data.	IEEE transactio n	Survey	The survey has performed the security analysis over EnCoRe system and suggested the preventive measures for it
Essa et al. [12]	The Hadoop based parallel and distribute big data analysis	IEEE transactio n	Hadoop-based data analysis	The method outcomes with the efficient results and solves the issues related in Hadoop
Cau et al. [13]	The detection method of service Denial based on Entropy for the cloud data centers.	IEEE transactio n	Entropy-based cloud computing	analyzed the attack of malicious over the virtual machines and developed the entropy based method to monitor the threats on the virtual machine and outcome with the significant data center results
Zheng et al. [14]	the internet of vehicles based method of big data analysis over the cloud computing	IEEE transactio n	Internet of vehicles	The authors have surveyed the issues in big data incorporated with the IOV and improve the IOV capability
Gohil et al. [15]	the performance analysis factors of the MapReduce applications in cloud-based Hadoop for big data	IEEE transactio n	Hadoop	The conclusion of the authors work states that the work decreased the execution time and enhanced the performance

5. RESEARCH GAP OF EXISTING WORK

From the survey analysis of the recent researches, the security, and privacy in big data is not on the mark. No proper work has been done to achieve 100% privacy and security. There needs a proper research to overcome the issues of real-time big data applications and privacy preservation.

6. FUTURE RESEARCH LINEUP

The future study can be extended as following way:

1. Dataset design parameters can be formulated, specializations of top down can be formulated, and the privacy preservation method can be established.
2. A framework can be designed for conducting computation over Top-Down Specialization based on parallelization provisioned by Map Reduce on the cloud for effective data anonymization.
3. Optimal privacy is achieved by adopting Hadoop to ensure the privacy preservation in complete communication traffic.

7. CONCLUSION

The modern ethnology like cloud computing will offer a scalable service for the big data with optimized cost. But the concern of privacy and security is still unsolved. This paper discussed the survey over the cloud-based big data security and privacy preservation. The survey discussed the recent work carried for privacy preservation and also existing research gap. The significance of this survey paper It gives an idea to improve the security and privacy of big data.

8. REFERENCES

[1] M. Mayilvaganan and M. Sabitha, "A cloud-based architecture for Big-Data analytics in smart grid: A proposal," Computational Intelligence and Computing Research (ICCIC), IEEE International Conference, Enathi, pp. 1-4, 2013

[2] G. Baciu, C. Li, Y. Wang and X. Zhang, "Clouds: Cloud-based cognition for large streaming data," Cognitive Informatics & Cognitive Computing (ICCI*CC), 2015 IEEE 14th International Conference, Beijing, 2015, pp. 333-338.

[3] Y. Tamura, K. Miyaoka and S. Yamada, "Reliability analysis based on three-dimensional stochastic differential equation for big data on cloud computing," 2014 IEEE International Conference on Industrial Engineering and Engineering Management, Bandar Sunway, 2014, pp. 863-867.f

[4] Guo, Jian, Zhao-Meng Zhu, Xiu-Min Zhou, and Gong-Xuan Zhang. "An instances placement algorithm based on disk i/o load for big data in private cloud." In Wavelet Active Media Technology and Information Processing (ICWAMTIP), 2012 International Conference on, pp. 287-290. IEEE, 2012.

[5] Pant, Vinay Kumar, Jyoti Prakash, and Amit Asthana. "Three step data security model for cloud computing based on RSA and steganography." Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on. IEEE, 2015.

[6] S. Luo; G. Zhang; C. Wu; S. Khan; K. Li, "Boafft: Distributed Deduplication for Big Data Storage in the Cloud," in IEEE Transactions on Cloud Computing , vol.PP, no.99, pp.1-1.

[7] Lemoudden, Mouad, and Bouabid El Ouahidi. "Managing cloud-generated logs using big data technologies." Wireless Networks and Mobile Communications (WINCOM), 2015 International Conference on. IEEE, 2015.

[8] Adnan, Muhammad, et al. "Minimizing big data problems using cloud computing based on Hadoop architecture." 2014 11th Annual High Capacity Optical Networks and Emerging/Enabling Technologies (Photonics for Energy). IEEE, 2014.

[9] Banditwattanawong, Thepparit, Masawee Masdisornchote, and Putchong Uthayopas. "Economical and efficient big data sharing with i-Cloud." 2014 International Conference on Big Data and Smart Computing (BIGCOMP). IEEE, 2014.

[10] Spicuglia, Sebastiano, et al. "Optimizing capacity allocation for big data applications in cloud datacenters." 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). IEEE, 2015.

[11] Sharif, Ather, et al. "Current security threats and prevention measures relating to cloud services, Hadoop concurrent processing, and big data." Big Data (Big Data), 2015 IEEE International Conference on. IEEE, 2015.

[12] Essa, Youssef M., Gamal Attiya, and Ayman El-Sayed. "Mobile agent based new framework for improving big data analysis." Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on. IEEE, 2013.

[13] Cao, Jiuxin, et al. "Entropy-based denial-of-service attack detection in cloud data center." Concurrency and Computation: Practice and Experience 27.18 (2015): 5623-5639.

[14] Zheng, Di, Kerong Ben, and Hongliang Yuan. "Research of Big Data Space-Time Analytics for Clouding Based Contexts-Aware IOV Applications." Advanced Cloud and Big Data (CBD), 2014 Second International Conference on. IEEE, 2014.

[15] Gohil, Parth, Dweepna Garg, and Bakul Panchal. "A performance analysis of MapReduce applications on big data in cloud based Hadoop." Information Communication and Embedded Systems (ICICES), 2014 International Conference on. IEEE, 2014.