ISSN 2278-3091
Volume 8, No.2, March - April 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse34822019.pdf
https://doi.org/10.30534/ijatcse/2019/34822019

Security Mechanisms leveraged to overcome the effects of Big Data characteristics

P Amarendra Reddy[1], O Ramesh[2]
[1]Associate Professor, Department of Information Technology, MLR Institute of Technology, Hyderabad, INDIA, amarpanyala88@gmail.com
[2]Assistant Professor, Department of Information Technology, MLR Institute of Technology, Hyderabad, INDIA, rameshmalyadri@gmail.com

ABSTRACT

Data is right now an emerged amidst the most critical and valuable assets for organizations in each area. So the world is going to produces 180 ZettaBytes of data in 2025 and the volume is the major issue. Big data market increases from 42 billion to 102 billion dollars till 2027 with increasing in the growth of science, healthcare, educational and research development in industry and other fields in Internet of Things and Cloud Computing. Now day's Big data security and protection of data plays a vital role. Different security and protection algorithms have developed with Big Data that are not liable to be catching the regular Big Data security challenges. We have discussed the important ideas and strategies for the Big Data Security protection challenges and recognize look into difficulties to be routed to accomplish complete answers for data security in the Big Data. Identified the fundamental issues related security challenges and solutions in a Big Data architecture. A complete framework analysis is done on the Big Data Security Challenges with different techniques to find the needs and solution on each.

Key words : Big Data, Hadoop, MapReduce, Cloud Computing, Information Security

1. INTRODUCTION

Data is generated in a large volumes of different forms (structured, semi structured and unstructured) dataset of petabytes or exabytes. Big Data can provide big success opportunities. However, as with most emerging technologies, several characteristics are associated with big data problems that make them technically challenging [1].

By 2020 data generated at 1.7MB per second and increase to 44 ZB with impact of IoT devices. Goolge creates data 1.3 trillions of data. Face book generates 31.35 million. With the Cloud computing 80% of data is going stored by 2026. Hadoop is going to increase by 58% BY 2020 and 75% of decision makers using Big data analytics. According to CACR 38.7% of Big Data technology increased in between 2015 and 2020 [1,2].

Gartner in 2001, defined 3Vs in 2001. The 3Vs are increasing Volume, Velocity and Variety. Gartner in 2012, updated V's clarify the essential characteristics of Big Data and it can't disregard. Big Data 42 characteristics to improve the performance and security [1,3].
The three principal challenges is approaching data security [4]:

1. Input Data
2. Data Storage
3. Output Data

Digital security in Big Data is a measure issue. Few companies like Thales (Vormetric), Cloudwick, Logtrust, IBM and Gemalto
Big Data not just expands the size of the difficulties identified with protection and security as they are tended to in customary security the board, yet in addition make new ones that should be drawn closer recently [5]. Big Data won't accomplish the required dimension of trusting on the 3V's characteristics.

1.1. KDD

In the following four ways data can be considered in the KDD in the figure.1 [6].
a. Handling: It chooses irregularities of missing data areas and expels them during the time of combing the data. Structure with the goal that it very well may be perused rapidly, to create and achieve potential outcomes.
b. Transformation: It moves data into proper structures for mining. The Data isn't introduced in its legitimate structure, and in this way, it must be dealt with to speak to a sort that can produce some valuable data.
c. Mining: Calculations are utilized to extricate data.
d. Pattern Evaluation: Once the data is separated, designs are then assessed to get learning on patterns.

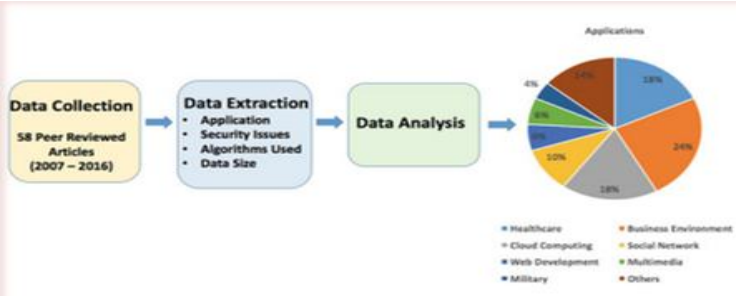

Figure 1: KDD analysis of data

1.2. Data Privacy

Consistently, a lot of data is produced and handled in a variety of businesses [2,7,8]. Along these lines, the protection of data

can be built up by efficient systems. In the table.1. shown the difficulties to security insurances [6].

Table.1: Different difficulties to security insurance

Data Providing	Data Collecting	Data Mining	Decision Making
Supply data according to the need of collectors	Retriever data from providers of client's daily needs or sensitive data. Before this data send to on internet this stage has to be done.	Its algorithms gets data from collecting step but risky in confidentiality [10]. Private data leaking.	Gets data from data miners. Confidentiality rules have to maintain along with authority, accuracy, objectivity, currency and coverage [10].
[10] Encryption tools like Cypher Text and AdBlocker are used.	Over comes the relationship between semi identifier and properties	Over comes with Probabilistic distortion	PolicyMakers and Competitors get affected if output revealed.

2. BIG DATA TECHNOLOGIES

In the Big Data the management, security [11]and governess show in the figure.2 based on Hadoop Ecosystem with different tools and Map-Reduce processing [13] is shown in the figure.3 and figure.4 [14].

Figure 2 : Typical Big Data architecture

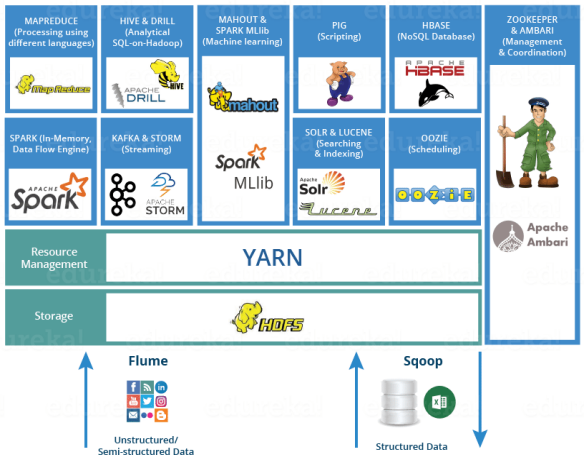

Figure.3: Hadoop Ecosystem

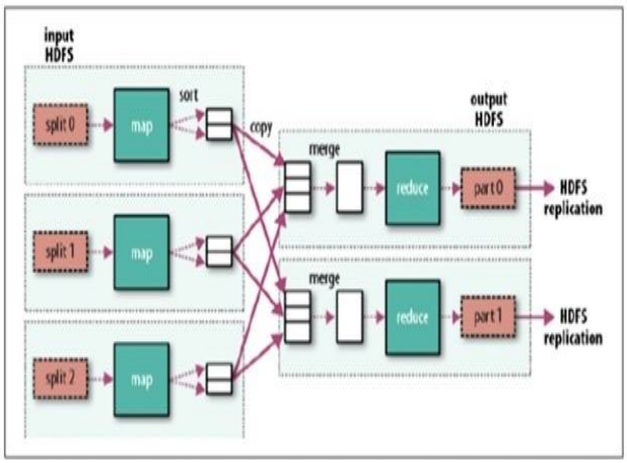

Figure 4: Map-Reduce Processing

2.1. HDFS and MapReduce

In the table.2 clearly represents the difference between in HDFS and Map Reduce components and its features .

Table.2: Components HDFS and Map Reduce

Type	Description	Components		Features
HDFS (Hadoop Distributed File System)	Primary Data Storage.	Name Node *(Master)*	Stores Metadata (**blocks**, location, Rack) runs file system [1]	Robust. Cost Effective. Faster Processing. Flexibility. Fault Tolerance.
		Date Node (Slave)	**Read-Write operations**. Replica block. Apply Handshaking principle. **Operations:** Block replica creation, deletion, and replication [2]	
Map Reduce	Provides data processing. (**key-value pairs**)	Map phase	Data splits into Key-Value pair	Simplicity. Scalability. Speed. Fault Tolerance.
		Reduce phase	Mapper gives the Output. Reducer combines and changes those input based on the key[3].	

Big Data situations don't generally organize security [4,15], and it was hence that we chosen to complete research so as to find the primary security challenges as for Big Data, alongside the arrangements, strategies, or methods proposed by scientists in order to accomplish security in Big Data frameworks. In the table.3 represents the security challenges and its solutions.

Table 3: Security Challenges and Solutions

Security Challenges of Big Data	Solution
Real-Time Monitoring: It stores the no. of security alerts raised, identifies the drawbacks and tracks the threats [5]	**Layered Protection:** Increasing the both in and out layers [11]
Granular Audits: In the live streaming it cannot detects the attacks due to this auditing is required [12]	**Protection of Different Domains:** Distributed system is needed to handle security[13]
Secure Computations in Distributed Systems: Parallel-Computing with storage[14].	**Hierarchical Protection:** Original data in different organization to provide access
Secure Data Storage and Transactions-Logs: Data moving into different layers not a big issues but the amount output data during the transactions is to be stored [15].	**Time-Sharing Protection:** Data during the sharing.
Endpoint Evaluation: Input data is to be validate from others.	**3KDEC Algorithm:** Symmetric Key-Block Algorithms is need [15].

3. BIG DATA SECURITY ALGORITHMS

In the table.4 represents the different security algorithms used and its description [13,14,15].

Table 4: Security Algorithms and Description

Algorithm/Key terms	Description
MapReduce [4]	It is key-value based program
AES : Advanced Encryption Standard) [13]	Symmetric Block-Cipher algorithm with block size: 128(192,256) bits and key size: 128 bits. Rounds 10(or 12, 14) for encryption.
Multilevel identity encryption [14]	Attributes based encryption
ORAM : Oblivious Random Access Memory.	Clients with small secure memory storage. To protects from caches attacks
XACML : eXtensible Access Control Markup Language.	Implements XML for access control
Data masking	It provides data breaches, loss, service, insecure and malicious
Random 4 [14]	It provides SQL Injection encryption
StarLight [15]	It is used to alert the persons on Sea Port areas
3DES : Data Encryption Standard [16]	It is a Symmetric Block-Cipher, faster than AES and keysize is 112 or 168 bits. Block size 64 bits. More secure than DES
VPN : Virtual Private Network [13,15]	It provide security among the threats
OBEX : OBjective EXchange [25]	Exchange Binary-data with other devices
MuteDB : Multi-User relaTional Encrypted DataBase [26]	It is used in cloud database for confidentiality

4. BIG DATA SECURITY CHALLENGES FRAMEWORK ANALYSIS

According to Cloud Security Alliance (CSA) organization the big data security challenges in the figure.5 and into four groups [5,27,28,29,30].

1. Infrastructure Security
2. Data Privacy
3. Data Management and Integrity
4. Integrity and Reactive Security

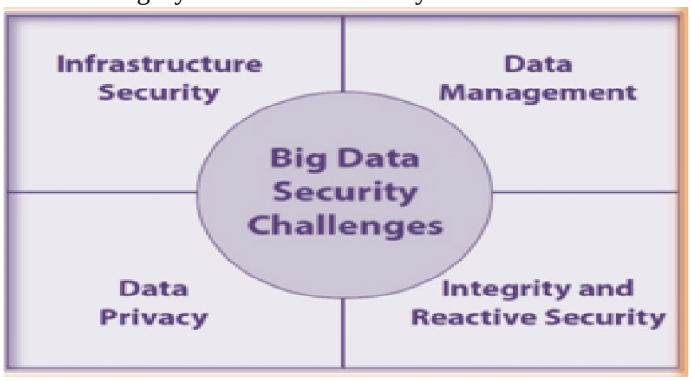

Figure.5: Security Challenges

The framework challenges are again divided into sub groups according to its impelemntation in the figure.6 . The architecure of the Big Data Challenges with data source with IoT devices [16,17] and cloud processing [12] and output to external devices in the figure.7.

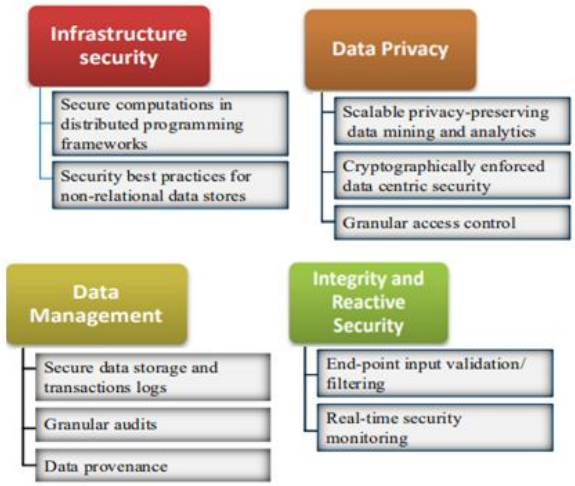

Figure 6: Big Data Security Challenges Framework Analysis

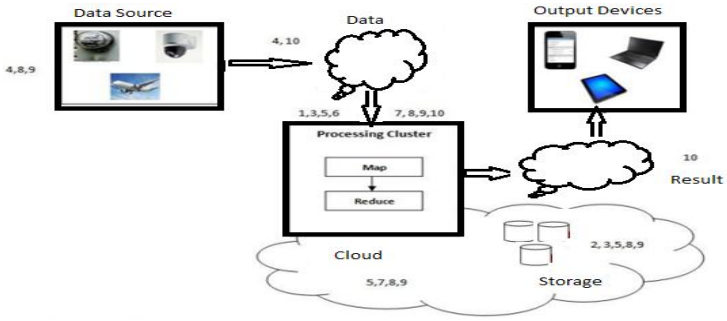

Figure.7: Big Data Security Challenges Architecture

Key security and privacy threats regularly require the three particular issues[13,17,18,19,26]:

1. Modeling: A risky standard that covers the majority of cyber attack [32].

2. Analysis: Reports the arrangements dependent on the threat display.

3. Implementation: Applying the results in the actual framework [33].

5. ANALYSIS OF RESULTS

In the figure.8 clearly analysis the Big Data security challenges with key factors with

Security principles [23,26,27,30,31,32]. In the Table.4. represented the infrastructure security techniques, needs and with its solutions. In the Table.5. represented the data privacy security techniques, needs and with its solutions, Table.6. represented the Data Management techniques, needs and solutions. Table.7. represented the Integrity and Reactive Security techniques, needs and solutions. Figure 10. represented the Papers grouped by main categories the Big Data Security Challenges with percentage of usage. Figure 11. Comparison of Big Data Security Challenges with framework analysis.

Figure.8: Big Data Security Challenges Analysis

Table.4: Infrastructure security techniques, needs and solutions

Infrastructure Security		
Techniques	**Needs**	**Solutions**
Hadoop Security [13]	It needs Authenticity.	G-Hadoop [18], SecureAccessSystem,
		New encryption techniques with new schema
Availability	Extension of a Hadoop implementation .	More active Name Nodes at a time to reduce the fault-tolerance [19].
Security for Architecture	Architecture security for data with Authenticity, Availability and Integrity [16].	Hadoop file system[3].
Authentication	The results of the data.	Signcryption Scheme for Identification [22]
Communication Security[23]	Big Data Ecosystem.	Security techniques for data transferring [20].

Table 5: Data privacy security techniques, needs and solutions

Data privacy		
Techniques	**Need**	**Solutions**
Cryptography [16]	Securing Data privacy. To protect data for limited time. Traditional techniques not useful.	Bitmap Encryption Scheme: User's Privacy. PigLatin: To analysis and program transformation [13,14]
Access Control	Restricting the unauthorized user access	MapReduce: key-value level Framework : Access Control
Confidentiality	Data is only accessed by authorized user with encryption and decryption method	Computing on Masked Data (CMD). Trusted Scheme for Hadoop Cluster (TSHC) [21,22]
Privacy for Social Networks	Protects personal data with high secure techniques.[23]	Protection of data privacy
Privacy-Preserving Queries	With encryption and decryption method used for modification data	High level encryption and decryption.[24]

Privacy for Differential	Reduce the identities access of users from unauthorized users	Encrypting the data with noise or private key [16]
Anonymisation [28]	It shares the data by hiding confidential data	Top-Down Specialization (TDS). Bottom-Up Generalization (BUG) [28]

Table.6: Data Management techniques, needs and solutions

Data Management		
Techniques	**Need**	**Solutions**
Security at Collection or Storage	A huge amount of data. to protect data owners' privacy	Privacy for Acceptance and Security storage [29]
Policies for Government or laws	Non-repudiation: Assurances of denying	Protectusers privacy [30]
Sharing Algorithms	Providing the security between the transaction with high privacy	Data sharing. Nested Sparse sampling and co-prime sampling [31].

Table 7: Integrity and Reactive Security techniques, needs and solutions

Integrity and Reactive Security		
Techniques	**Need**	**Solutions**
Integrity	Assurances for the data in the same form in the transmit i.e., no alteration	MapReduce: Identifying the unauthorized users [18,19.20]
Attack Detection	Early threat detection reduces the risk of accessing the data	MapReduce: Intrusion Detection System(IDS) [22,26]
Recovery	Develop the system from data backups and disasters	Disaster recovery system [27]

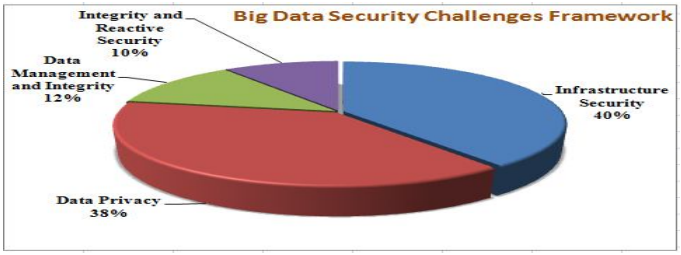

Figure 10: Papers grouped by main categories.

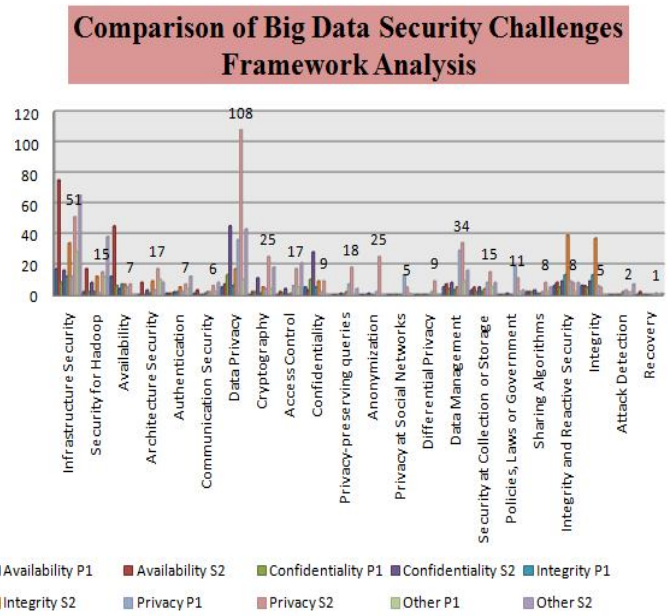

Figure 11: Comparison of Big Data Security Challenges with framework analysis.

6. CONCLUSION

Now day's Big data security and protection of data plays a vital role. Different security and protection algorithms have developed with Big Data that are not liable to be catching the regular Big Data security challenges. We have discussed the important ideas and strategies for the Big Data Security protection challenges. Identified the fundamental issues related to security challenges and solutions in Big Data Architecture. A new framework analysis is done on the Big Data Security Challenges with different techniques to find the needs and solution on each.

ACKNOWLEDGEMENT

The authors would like to acknowledge Science and Engineering Research Board, India for financial support..

REFERENCES

1. ShirishaNalla and Sheikh Gouse, KVD Kiran. (2018). Machine Learning Challenges Of Big Data International Journal of Pure and Applied Mathematics Volume 120 No. 6 2018, 3377-3386
2. Dr P Amarendra Reddy, Dr Sheikh Gouse,Dr P

Bhaskara Reddy. (2018) . Big Data 42 Characteristics to Improve the Performance and Security, JARDCS vol. 10 Issues-7, 2018 Pages 1524-1537. ISSN 1943-023X

3. Kumar, J. Pradeep, Sheikh Gouse, and P. Amarendra Reddy. Migration of Big Data Analysis from Hadoop's MapReduce to Spark. First International Conference on Artificial Intelligence and Cognitive Computing. Springer, Singapore, 2019.
4. Tankard, Colin. Big data security. Network security 2012.7 (2012): 5-8. https://doi.org/10.1016/S1353-4858(12)70063-6
5. Suthaharan, Shan. Big data classification: Problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS Performance Evaluation Review 41.4 (2014): 70-73. https://doi.org/10.1145/2627534.2627557
6. Sheikh Gouse, P Amarendar Reddy, P Bhaskara Reddy. (2014) Basic Data Mining Models and Tools Data Mining International Journal of Emerging Trends in Engineering and Development (IJETED) ISSN 2249-6149 Issue 4, Vol.2 (March 2014) & pages 804 -817
7. Dr. Sheikh Gouse, RaswithaBandi. (2018) Big Data Technique To Provide Security And Simplifying The Heterogeneous Data Indian Journal of Scientific Research (IJSR). 17(2): 432-437, 2018
8. Dr.Sheikh Gouse, S.Sravya, G.Charitha, Harish sai. (2018) Working with employee Datasets using Hive Complex Data types International Journal of Research in Electronics and Computer Engineering (Ijrece) Vol. 6 Issue 2 Apr.-June 2018.Pages 779 -782. ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE)
9. Dr. Sheikh Gouse,RubiyaSubair, .(2018) MAA Access Control Over Cloud Server Using AAC Scheme in Cloud Computing, Journal of Advanced Research in Dynamical and Control Systems(JARDCS) vol.10 Issues-5, 2018 Pages 781 - 789. ISSN 1943-023X
10. Dr. Sheikh Gouse, Raswitha Bandi, Dr.P.Amrendar Reddy. (2018) A Comprehensive Survey On Big Data Analytics And Tools, International Journal of Computer Engineering and Applications,Volume XII, Issue I, Jan. 18, www.ijcea.com ISSN 2321-3469.
11. Dr. Sheikh Gouse, B. Madhuravani, B. Swapna. (2017), Improved Network Lifetime In Wireless Sensor Networks Using Eliptic Curve Cryptography, International Journal of Mechanical Engineering & Technology (IJMET) July 2017 Volume.8, Issue:7, Pages:308-318
12. B.Madhuravani,Dr. P. Bhaskara Reddy, Dr.SheikhGouse, Swapna Bhumandala (2017), Secure Authentication And Dynamic Encryption Using ECC And Wireless Network, Journal of Advanced Research in Dynamical and Control Systems(JARDCS) 2017 Issues-12, Page1131-1144.
13. RaswithaBandi,Dr. Sheikh Gouse, Dr.J.Amudhvel, (2017), A Comparative Analysis For Big Data Challenges And Big Data Issues Using Information Security Encryption Techniques, International Journal of Pure and Applied Mathematics, 2017 Vol.8 Pages183-189 ISSN 978-3-319-63644-3
14. Dr. Sheikh Gouse, Y. SudhaLaxmi, A. Mahendar. (2017), Customer Complaint Analysis Using Hadoop (Consumer Analysis), International Journal Of Engineering And Computer Science, 6(4). ISSN:2319-7242 Volume 6 Issue 4 April 2017, Page No. 21112-21116
15. Nirosha, K., B. Durgasree, and Sheikh Gouse. iHome: Bio-Health Intelligent Mobile System UsingIloT. International Journal of Innovations in Engineering and Technology (IJIET)7.4 (2016).
16. Koluguri, Abhishek, Sheikh Gouse, and P. Bhaskara Reddy. Text steganography methods and its tools. International Journal of Advanced Scientific and Technical Research 2.4 (2014): 888-902.
17. Nirosha, K., B. Durga Sri, and Sheikh Gouse. Smart Heartbeat Monitoring System Using Machine Learning. First International Conference on Artificial Intelligence and Cognitive Computing. Springer, Singapore, 2019 https://doi.org/10.1007/978-981-13-1580-0_35
18. Zhao, J.; Wang, L.; Tao, J.; Chen, J.; Sun, W.; Ranjan, R.; Kołodziej, J.; Streit, A.; Georgakopoulos, D. A security framework in G-Hadoop for big data computing across distributed Cloud data centers. *J. Comput. Syst. Sci.* 2014, *80*, 994–1007. https://doi.org/10.1016/j.jcss.2014.02.006
19. Cohen, J.C.; Acharya, S. Towards a trusted HDFS storage platform: Mitigating threats to Hadoop infrastructures using hardware-accelerated encryption with TPM-rooted key protection. *J. Inf. Secur. Appl.* 2014, *19*, 224–244. https://doi.org/10.1016/j.jisa.2014.03.003
20. Ma, Y.; Zhou, Y.; Yu, Y.; Peng, C.; Wang, Z.; Du, S. A Novel Approach for Improving Security and Storage Efficiency on HDFS. *Procedia Comput. Sci.* 2015, *52*, 631–635. https://doi.org/10.1016/j.procs.2015.05.062
21. He, S.; Wu, Q.; Qin, B.; Liu, J.; Li, Y. Efficient group key management for secure big data in predictable large-scale networks. *Concurr. Comput.* 2016, *28*, 1174–1192. https://doi.org/10.1002/cpe.3574
22. Wei, G.; Shao, J.; Xiang, Y.; Zhu, P.; Lu, R. Obtain confidentiality or/and authenticity in Big Data by ID-based generalized signcryption. *Inf. Sci.* 2015, *318*, 111–122. https://doi.org/10.1016/j.ins.2014.05.034
23. Colombo, P.; Ferrari, E. Privacy Aware Access Control for Big Data: A Research Roadmap. *Big Data Res.*2015, *2*, 145–154. https://doi.org/10.1016/j.bdr.2015.08.001
24. Ulusoy, H.; Colombo, P.; Ferrari, E.; Kantarcioglu, M.; Pattuk, E. GuardMR: Fine-grained Security Policy Enforcement for MapReduce Systems. In Proceedings of the 10th ACM Symposium on Data, Computer and Communications Security, Singapore, 14–17 April

2015; pp. 285–296.
https://doi.org/10.1145/2714576.2714624

25. Kepner, J.; Gadepally, V.; Michaleas, P.; Schear, N.; Varia, M.; Yerukhimovich, A.; Cunningham, R.K. Computing on masked data: A high performance method for improving big data veracity. In Proceedings of the 2014 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 9–11 September 2014; pp. 1–6.
https://doi.org/10.1109/HPEC.2014.7040946
26. Quan, Z.; Xiao, D.; Wu, D.; Tang, C.; Rong, C. TSHC: Trusted Scheme for Hadoop Cluster. In Proceedings of the 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), Xi'an, China, 9–11 September 2013; pp. 344–349.
https://doi.org/10.1109/EIDWT.2013.66
27. Kuzu, M.; Islam, M.S.; Kantarcioglu, M. Distributed Search over Encrypted Big Data. In Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, San Antonio, TX, USA, 2–4 March 2015; pp. 271–278.
https://doi.org/10.1145/2699026.2699116
28. Irudayasamy, A.; Arockiam, L. Scalable multidimensional anonymization algorithm over big data using map reduce on public cloud. *J. Theor. Appl. Inf. Technol.* 2015, *74*, 221–231.
29. Weber, A.S. Suggested legal framework for student data privacy in the age of big data and smart devices. In *Smart Digital Futures*; IOS Press: Washington, DC, USA, 2014; Volume 262.
30. Wang, Y.; Wei, J.; Srivatsa, M.; Duan, Y.; Du, W. IntegrityMR: Integrity assurance framework for big data analytics and management applications. In Proceedings of the 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA, 6–9 October 2013; pp. 33–40.
https://doi.org/10.1109/BigData.2013.6691780
31. Liao, C.;Squicciarini, A. Towards provenance-based anomaly detection in MapReduce. In Proceedings of the 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Shenzhen, China, 4–7 May 2015; pp. 647–656.
https://doi.org/10.1109/CCGrid.2015.16
32. Tan, Z.; Nagar, U.T.; He, X.; Nanda, P.; Liu, R.P.; Wang, S.; Hu, J. Enhancing big data security with collaborative intrusion detection. *IEEE Cloud* Comput. 2014, 1, 27–33.
https://doi.org/10.1109/MCC.2014.53
33. Prasad TV, S. K Kumar, Ajay Kumar, Ch Uma Devi, 5B Nanda Kishore," A Novel Approach of De duplication of Records using Febrl Algorithm and Data Mining", International Journal of Advanced Trends in Computer Science and Engineering, Advanced Trends in Computer Science and Engineering.