BIG DATA AND BIG DATA ANALYTICS: A LITERATURE-BASED RESEARCH ON PRIVACY AND SECURITY -A DECADE TOUR

[1]Sule, A. A., [2]Okunoye, O. B., [1]Oyefolahan I. O.

[1]School of Information & Communication Technology
Federal University of Technology
Minna, Nigeria.

[2]Department of Computer Science,
University of Lagos,
Lagos, Nigeria.

suleaishat990@gmail.com, bukunoye@unilag.edu.ng, o.ishaq@futminna.edu.ng.

Abstract *— Over time, there has been an increase in the the world's interest in big data and analytics as data forms a major asset of every operable organization today. However, with the high volume of data generated at regular intervals, security and privacy in all dimensions of big data becomes a major concern. Since its inception, numerous literatures have been researched in the field of big data analytics security and privacy. Therefore, a study is required to provide a proper understanding of the security and privacy requirements of Big Data Analytics and ensure that the full potential of data is exploited.*

This paper seeks to perform a literature-based research on privacy and security with respect to big data and big data analytics over the past 10 years by visiting several authoritative journals for articles related to big data and analytics' privacy and security. The study gives a comprehensive review of big data and its related security and privacy concerns. At the end, it was discovered that although so much has been done to provide the needed security to data, more work still needs to be done to cope with the technological advancements and the rapid increase in data generation from different computing environment.

Keywords: *Big Data; Big Data Analytics; Security; Privacy.*

1.0 INTRODUCTION

The complexity of data is on the increase owing to high level of computing and internet growth. Big data has been described majorly by its volume which is attributed to the rising amount of data generated through the Internet of Things(IoT), scientific simulations, Next-Generation Sequencing (NGS) and other sources of data (Gholami & Laure, 2016). As the name implies, Big data refers to large size files. Although the four V's (Volume, Velocity, Variety, Veracity) describe big data, the volume was the essence of its existence. In recent times, another V (Value) was added to the existing four dimensions of big data making it five V's (Liu, 2016)

Big data's volume characteristics, which happens to be the most obvious attribute, suggest that protection of data is sacrosanct as such requires more effort to protecting the large amount of regularly generated data. Measures have been put in place since the advent of big data to ensure data privacy and security.

As data continues to grow in speed and variety, analysis becomes necessary in order to convert the data into meaningful information. Big Data Analysis is the use of statistical techniques to obtain information, which can then be used to improve processes or to help in decision making (Runkler, 2016). Therefore, numerous analytics tools were developed and used over the years, the most common of which is the Hadoop, which consists of two components–the storage part–Hadoop Distributed File System (HDFS) and the processing part–MapReduce. (Gautam & Rana, 2018).

The world is generating zetta bytes of data on a daily basis. Due to the large nature of the data, analysis becomes more demanding and security becomes more difficult and a major concern. Big data analytics includes many security issues, including real-time data analysis secure storage of transaction records and data, granular access control and data provenance (Jayasingh et al, 2016). Security breaches is no doubt a big deal for big data as the data is coming from different sources and of different varieties at different speed.

In this paper, we present an overview of big data and big data analytics as well as privacy and security concern in big data analytics. Furthermore, a review of the techniques adopted so far to ensure the privacy and security of big data analytics was also portrayed. The study therefore gives an insight into some security and privacy challenges evident in big data.

2.0 LITERATURE REVIEW

2.1 Big data

Big data is a broad area and has caught the interest of many researchers in recent times. There are several definitions to big data but all have one thing in common and that is the fact that big data encompasses huge amount of data. Although no standard definition exists for big data, it has been described by several researches as follows:

- Jacobs (2009) in his research described big data as *"data that is too large to be placed in a relational database and analysed with the help of a desktop statistic/visualization package-data, perhaps, whose analysis requires massively parallel software running on tens, hundreds, or even thousands of servers."*
- Rouse (2011) described Big Data as *"description of the voluminous amount of unstructured and semi-structured data a company creates or data that would take too much time and cost too much money to load into a relational database for analysis"*
- Lohr (2012) defined big data as *"A meme and a marketing term, for sure, but also shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making decisions.".*
- IDC (2013) gave the definition of big data as having *"three main characteristics: the data itself, the analytics of the data, and the presentation of the results of the analytics. Then there are the products and services that can be wrapped around one or all of these Big Data elements."*
- Patil & Seshadri (2014) *also gave their own description of big data as: "the agglomeration of large and complex data sets, which exceeds existing computational, storage and communication capabilities of conventional methods or systems.*
- Venkata 2015 described big data as *"large amount of data which requires new technologies and architectures to make possible to extract value from it by capturing and analysis process.*
- Barznji & Atanassov (2016) simply defined big data as "*a large volume, heterogeneous, distributed data.*
- Mills (2017) described big data as "*as digitally encoded information of unprecedented scope or scale about a phenomenon, which has relationality with other networked data."*
- Gautam & Rana (2018) in their own definition of big data said: "*Big data is a term used to describe the exponential growth and availability of data, having structured, unstructured and semi-structured data, whose size (volume), complexity (variability), and rate of growth (velocity) make them difficult or even impossible to be managed and analyzed using conventional software tools and technologies."*

Having reviewed the different definitions of big data, we can say categorically that big data is described first and foremost by the size which is what motivated the name 'big data' although discussion on big data emerged recently that "big" is no longer the defining characteristic, but, rather, how "smart" the data is -which means the meaningful information that can be deduced from the data (George, 2014).

2.1.1 Dimensions of Big data

a. Volume: This defines the amount of data being generated which is usually in the range of terabytes to zettabytes.

b. Velocity: This describes the rate at which data is being generated. Today, the availability of applications such as smart cities and smart planet contributes to data being generated at very high frequency which translate to high speed data streams. Meanwhile, it is not only how quickly data arrives but how quick

it can be processed. Speed of processing should be faster than the arrival of new data.

c. Variety Data today can come in many file format and can be categorized as (Structured, Semi- Structured & Un-Structured) on that basis.

1) Structured Set of information with a specified model for rehashing. It initially relies on a show of information Information model of type of business information that will be registered and how it will be processed, planned or surpassed, such an example makes storing, interpreting and manipulating the data less challenging for any system. (Jiang & Zheng, 2014). This contains RDBMS, Flat files in record format, multi-dimensional DBS.

2)Semi-Structured Collection of data not aligned with a standard predefined format. (Zheng & Jiang, 2014) Several data sources include emails, XML, CSV, TSV, etc.

3)Un-Structured Data refers to non-predefined patterned information Unstructured data is essentially message overwhelming yet may include date, numbers, and this information also has inconsistencies that make it difficult to comprehend custom PC programs as a contrast to structured information. (Zheng & Jiang, 2014) Many unstructured data sources are files, images, audio, video, etc.

d. Value: Various data's economic value varies significantly. Usually, a larger body of non-traditional data masks good information; the task is to determine what is important and then process and extract data for analysis Industrial processes as well as academic research have shown the great value of big data if correctly used (Jamiy et al, 2015).

e. Veracity: Ensuring the accuracy of data has become an issue as the size increases exponentially and that explains why the fifth V was included in the initial four V's of big data. The increased data size makes it difficult to decide the number of data can be trusted when making a major decision on such large volumes produced at a very high speed. Siewert, (2013) disclosed how hard it could be to detect a compromised source from one of thousands of security cameras, each generating thousands of video frames per hour.

2.1.2 Sources of Big Data

a. Traditional enterprise data - includes transactional ERP data, customer information from CRM systems, general ledger data, and web store transactions.

b. Machine-generated Isensor data - includes Call Detail Records ("CDR"), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), and trading systems data.

c. Social data - includes customer feedback streams, microblogging sites like Twitter, and social media platforms like Face book

2.1.3 Classification of Big Data

Big data is classified into different categories to understand better their characteristics according to Hashem el al (2015), the classification was done based on five aspects:

a. Data Sources

b. Content Format

c. Data stores

d. Data Staging

e. Data processing

Each of these categories and characteristics are described in Figure 1:

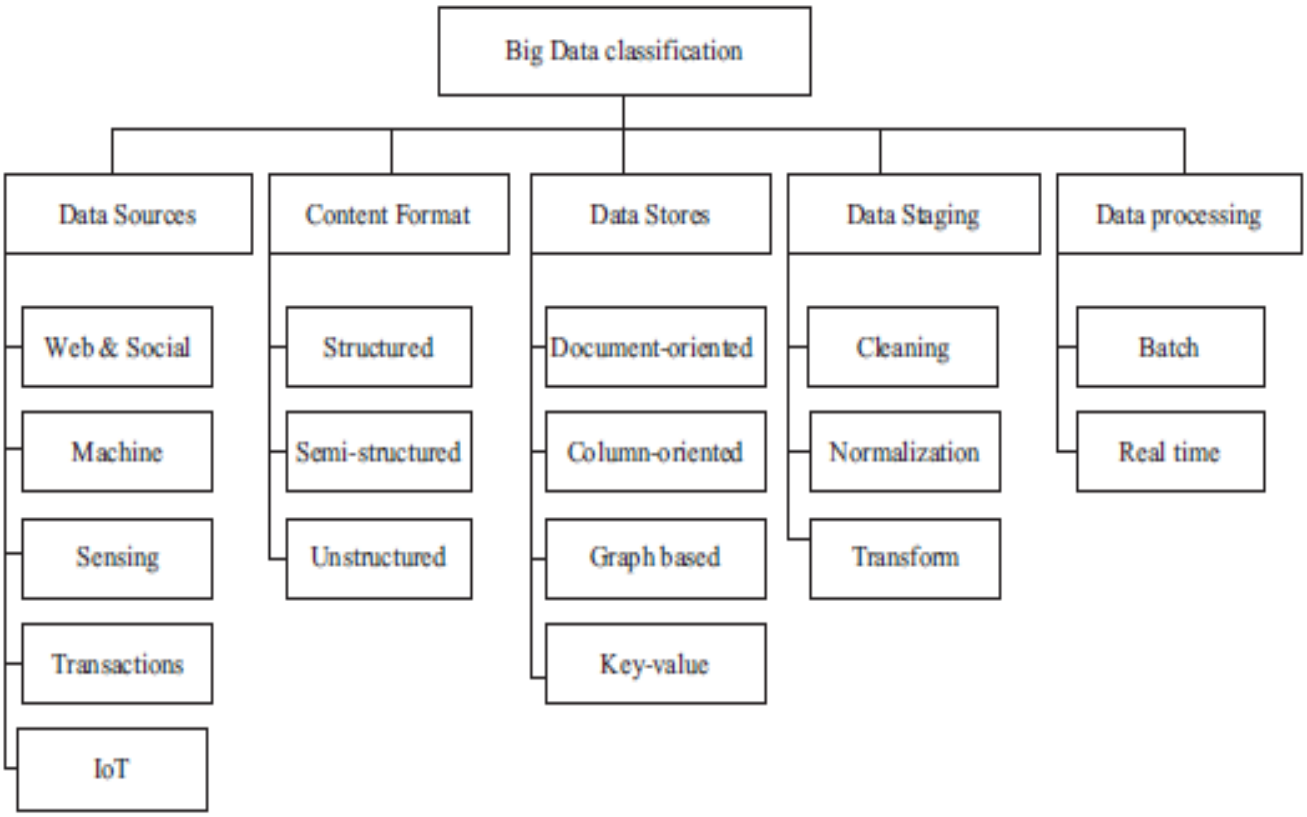

Figure 1: Classification of Big Data (Hashem et al, 2015)

2.1.4 Importance/Benefits of Big Data

The importance of big data cannot be overemphasized, as almost every interaction between businesses, individuals and governments is driven by data (Polonetsky & Tene, 2013); Companies and organization have shifted towards big data because of its ability to hasten decision making to enable them stay ahead of the game in any competitive environment. Big data is an inevitable tool for creating enormous value for the global economy, productivity, efficiency, growth and driving innovation.

Implementation of big data also provide organizations with a thorough understanding of their businesses which can lead to greater innovation and enhanced productivity (Singh, 2014). It is also a useful tool for gaining insights into individual or team behaviour and consumer preferences. Big data is notable in the generation of value for businesses although Wamba et al (2015) opined that there is inadequate empirical research to gain insights into values provided by big data to businesses. National policies like information infrastructure, optimization of natural resources and disaster alarming system are also being influenced by big data (Chih-Liang, 2015).

Big data is applicable in domains such as public administration, biochemistry, healthcare, retail and other interdisciplinary scientific researches

2.1.5 Big data Analytics

It is imperative to analyse data to obtain useful information as data in its raw form makes no impact on organizations that generate them. Data analytics as defined by is *"the application of mathematical methods to gain information from data, which can then be used to optimise processes or supports decisions"* (George, 2012). In analysing big data therefore, the typical statistical approach is not sufficient due the humongous amount of data (George, 2012). More complex techniques are required to gain useful insights from such data as large as big data; these techniques are but not limited to data mining, machine learning, cluster analysis, genetic algorithms and natural language processing (McKinsey Global Institute, 2011). The general architecture of big data

analytics as shown in Figure 2 indicates that big data analytics is comprised of three parts.

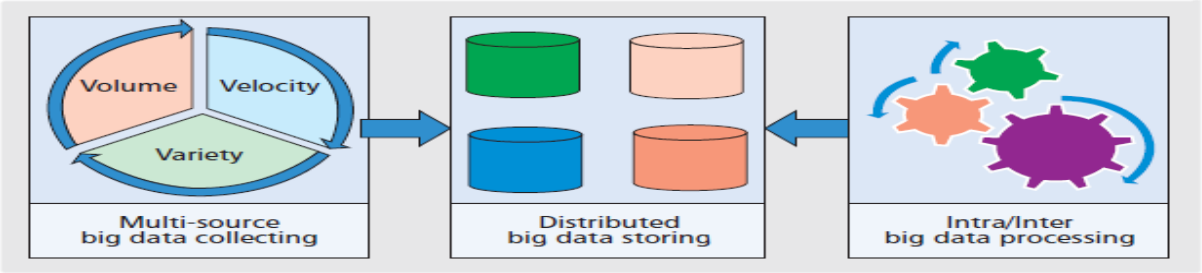

Figure 2. General architecture of big data analytics (Lu, et al, 2014).

Big data analytics is basically about two things –big data and analytics. Figure 3 shows how big data is transformed for analysis using any suitable analytics techniques.

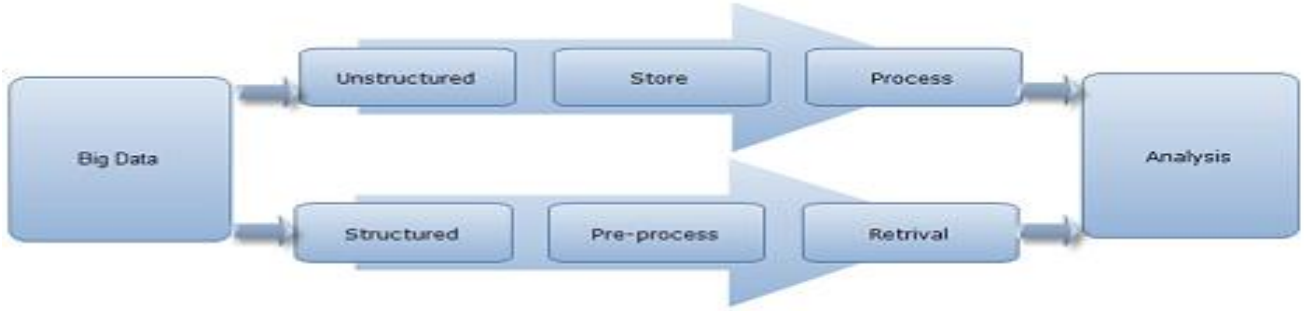

Figure 3. Transforming big data for analysis (Hashem et al., 2015)

2.1.6 Technologies for Big Data Analytics

Several technologies exist for analysing big data whether open-sourced or commercial. One of such open source technology for data analytics is the Hadoop Distributed File System (HDFS) (Singh, 2014) which is a long-term storage system for web logs.

a. Hadoop: Hadoop is an open-source software platform used with the MapReduce programming method for distributed storage and processing of big data. Hadoop modules are designed with a fundamental assumption that hardware faults are common occurrences and should be addressed by the system automatically. The core of Hadoop consists of two parts the storage part–Hadoop Distributed File System (HDFS) and processing part – MapReduce (Gautam & Rana, 2018).

b. ***Cloudera:*** Cloudera Inc. was founded in 2008 by experts from top companies such as Yahoo, Google etc. Cloudera was the leader in the development of personalized applications around Apache Hadoop core in terms of user base, it still has the top plays Although the core of its deployment is Apache Hadoop, it also has the Cloudera Management Suite as a proprietary product. This software helps to simplify Hadoop's implementation and includes a cluster control GUI framework.

c. Hortonworks: Hortonworks was launched in 2011 and has quickly joined Hadoop group's leading distributors. It provides a Core Apache Hadoop-based open source data platform for big data. Hortonworks is the only Hadoop provider to offer Apache Hadoop without any proprietary modules being included. Hortonworks HDP2.0 distribution can be downloaded directly from their website free of charge and can be easily installed. The Hortonworks team is responsible for many of the recent additions created by Hadoop, including Yarn, an enhanced version of basic MapReduce.

d. *MapR:* MapR is Apache Hadoop's open source version with several restrictions with their distributions, most of the Hadoop vendors tried to overcome those restrictions. All distributors have introduced their variety to the core open source component of Hadoop. MapR replaced the HDFS part with its own proprietary file system, known as the MapRFS. MapRFS helps add enterprise-grade capabilities to Hadoop, allowing more powerful data management and user-friendliness. Hadoop is introduced as a default feature of Ubuntu operating system by MapR through a new collaboration with the developer of Ubuntu operating system Canonical.

e. *IBM BigInsights:* IBM has a Spark and Hadoop suit for both cloud and on-site businesses Organizations would like to spend less time developing an enterprise ready Hadoop platform and getting more insights. A comprehensive solution is provided by IBM which includes Spark to quickly and easily scale analytics. Available on-premises, on-cloud, or incorporated with other systems currently in use.

2.2 Security and Privacy

Big data as applicable to areas such as social media, health care, social networking, and the Internet of Things (IoT) is connected with several security and privacy related problems. However, utilization of big data by organization requires that adequate security measures be established to protect data confidentiality, integrity and availability. Nevertheless, the issues with the security and privacy of big data for the aforementioned domains are highlighted in Table 1.

Table 1: Big Data Security and Privacy Issues

Domains	Security and Privacy Issues
Healthcare	Most healthcare data are certified by HIPPA, but this does not certainly mean that a patient's data is free of any security breach. A study conducted by the Ponemon Institute LLC (2012) showed that over the past two years, 94 percent of hospitals have experienced one or the other security breach, most of which were an internal rather than an external attack.
Social Networks	The large quantity of data generated daily from social networking is such that the data cannot be guaranteed utmost security and privacy. People share information on various platforms such as Twitter, Facebook, and so on; although some of these platforms such as Facebook provides users with options to set privacy but

	that is just to provide security at the user end but not at the end where the networking site is handled and developed. The daily increase in the popularity of social networking sites makes it prone to new treats that compromise the privacy and security of user data. However, the work of Cristin et al (2012) suggested that privacy bubbles be used as the actual boundary between users for picture sharing online.
Internet of Things (IoT)	IoT being the interconnection of several computing devices and people, can connect people and things to anything and anyone, anytime and anyplace using network. The security issues associated with IoT can therefore be tied to the fact that when people subscribe to online services that are free like the Emails and newsfeeds, they become source of data for the businesses. These data from online consumers are used by the online service providers for analysis and sometimes Outsourced and traded for further analysis to third parties without taking into account the online consumers (Perera, 2015). The era of IoT implies that more data will be generated as compared to time past owing to the availability of several wearable devices such as Apple Health Kit, Google Glass, Apple iWatch and Google fit that are capable of collecting sensitive information about users like financial status and health conditions (Sundmaeker et al (2010).
Social Media	A vast amount of media data in the form of videos, text, videos and audios are shared in various forms by people over several platforms and these trends is on the increase on a daily basis does not seem like there will be an end to it; hence, the rise in privacy and security concerns in the domain of social media

Security and privacy of big data and analytics cannot be over-emphasized especially being in a world where data is core in all business and individual environment. However, it has been said by Bertino & Sandhu (2005) that any data security solution that is comprehensive must meet the CIA (Confidentiality/secrecy, Integrity, Availability) concept. Confidentiality implies that the data must be free of unauthorized disclosure; Integrity implies data must be free of any form of modification; lastly, availability implies that data must be available to users of a system at all times irrespective of unexpected failure or error.

Whilst the concept of security is being confused with privacy, it is important to note that data privacy has additional requirements in comparison to data confidentiality which is key to achieving data privacy (Anton et al, 2007). These extra requirements include user consents for the use of personal data, compliance with privacy related regulations and deriving from the use of privacy-sensitive data. Bertino, 2015 mentioned that in any application domain we may think of, privacy is a critical factor to ensuring data safety.

A literature by Koutroumpis & Leiponen, (2013) identified big data sharing agreements as a privacy concern stating that the agreements level of big data sharing is poorly structured, informal, linked to isolated transactions and manually enforced. However, George (2014) proposed a solution to the big data-sharing problems recommending that data sharing agreements be linked into anonymization, data usage control, access control and other mechanisms for data protection and privacy. A similar research by Singh (2014), outlined ten (10) security challenges of big data which includes:

1. Secure computations in distributed programming frameworks
2. Security best practices for non-relational data stores
3. Secure data storage and transactions logs
4. End-point input validation/filtering
5. Real-time security/compliance monitoring
6. Scalable and compos able privacy-preserving data mining and analytics
7. Cryptographically enforced access control and secure communication
8. Granular access control
9. Granular audits
10. Data provenance

However, a machine learning model was developed to overcome the aforementioned security challenges in real-time big data analytics.

In a study carried out by Damiani (2015), security threat called big data leak which compromise the confidentiality and privacy of data was identified. He discussed how the Known, Detect, Contain and Recover paradigm could help contain risks associated with data disclosure with a recommendation that big data system design should not just focus on attack prevention but also recovering from attacks once detected.

A threat based on data mining is addressed by Dev et al (2012) in their study. According to them, this type of threat employs strategies and methods of data mining to collect data that are sensitive as well as valuable information. Another security risk to big data is data privacy and the disclosure of sensitive data that could affect individuals or organizations such as re-identification threat and wrong results threats (Jensen 2013).

Smith et al. (2012) examined personal privacy on social networks in their study and how social web users can control their data privacy. Kim et al. (2013) addressed the security of big data and provide a technique of security to efficiently harden and secure big data security by protecting the chosen attributes. Jensen (2013) addresses the problems of confidentiality in big data and how big data storage can be regulated under privacy complaint. Big data analysis can be used to boost security by collecting and analyzing information (structured or unstructured) from organizations (Cardenas et al. 2013).

Marchal et al. (2014) suggested a security model to analyze large amounts of data from a security point of view to track local business networks, execute network intrusion detection and protection activities, and conduct forensic analysis. In addition, another research focuses on the role of the consumer in protecting large data networks through the proposed security model (Xu et al. 2014). They argue in big data environment about four types of user role: data collector, data rovider, data miner, and decision maker (Xu et al. 2014). On the other hand, Big Data engineering is threatened by several threats and attacks Wu and Guo (2013) argue that the privacy and information assurance in big data is a major concern.

3. Methodology

The methods adopted here is depicted in figure 4 which involve the use of search engines such as Google, Google Scholar and Web crawlers to search for relevant papers relating to big data security and privacy using keywords such as big data, security and data privacy. While searching, the year of publication was also considered; since we are taking a 10 years' tour, it is imperative that we take into consideration the year of publication.

However, publications from 2009 to 2018 were considered. In addition, the papers that were considered are those published in reputable journals such as IEEE, Elsevier, Science Direct, etc.

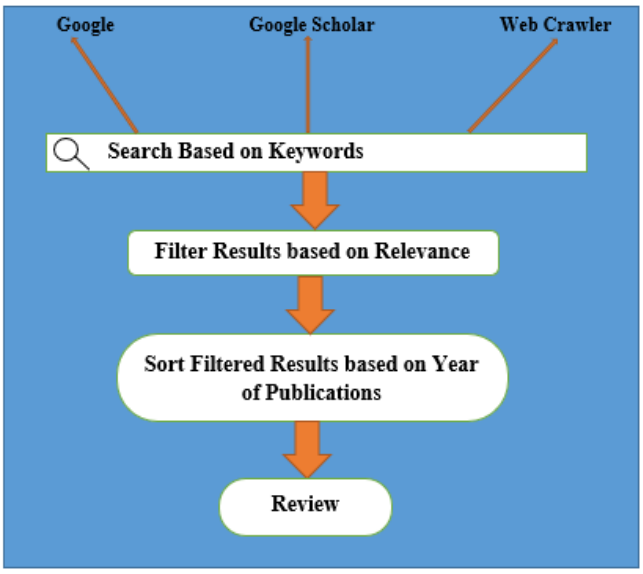

Figure 4: Adopted Methodology

Furthermore, the papers were sorted according to their year of publication with at least 10 papers in each year, some years had more than 20 papers like those published between 2013 and 2018.

4. Findings

Our findings from this study shows that new privacy threats are evident in big data owing to the advancement in computing technology and the rate at which data is being generated. Aside the voluminous nature of big data that brings about new security threats, easy accessibility to big data also pose a security challenge in the sense that big data is being easily accessed by organizations of all sizes through public cloud infrastructures made possible through the coupling of big data with public cloud environments as opposed to the past where it was only accessible to very large organizations such as governments and large enterprises that could afford the necessary infrastructures for hosting large amount of data. Thus, security and privacy of big data still needs a baby attention to protect user data from any form of security breaches.

5. CONCLUSION

Complexities arising from ever increasing data has called for security and privacy concerns and so it becomes imperative to put at the forefront of big data analytics measures to ensure data security and privacy. Security is an ongoing issue as such technological advancements would require new measures to be provided to cope with new threats that may compromise the confidentiality, integrity and availability of data. The challenges of big data

security and privacy is enormous as such, government should be more involved in big data security and privacy by constantly reviewing policies guiding user data collection and usage to cope with new antics of attackers. More so, analytics tools should be upgraded frequently to cater for new threats.

REFERENCES

Koutroumpis, Pantelis & Aija, Leiponen. (2013). Understanding the value of (big) data. 38-42. 10.1109/BigData.2013.6691691.

Runkler, T.A. (2016) Data Analytics: Models and Algorithms for Intelligent Data Analysis, 2nd ed.; Springer Vieweg: Wiesbaden, Germany; p. 1, ISBN 9783658140755.

Sam B. Siewert. 2013. Big data in the cloud – Data velocity, volume, variety and veracity. https://www.ibm.com/developerworks/library/bd-bigdatacloud/index.html.

Fatima EL Jamiy, Abderrahmane Daif, Mohamed Azouazi, and Abdelaziz Marzak. 2015. The potential and challenges of big data-Recommendation systems next level application. CoRR abs/1501.03424 (2015). http://arxiv.org/abs/1501.03424.

Yeh, Chih-Liang (2015): A qualitative change of personal information in the age of big data: A Study of court case in Taiwan, 26th European Regional Conference of the International Telecommunications Society (ITS), Madrid, Spain, 24-27 June 2015, International Telecommunications Society (ITS), Madrid

Smith, M., Szongott, C., Henne, B., and Voigt, G. Von. 2012. "Big Data Privacy Issues in Public Social Media," in *Proceeding of the Digital Ecosystems Technologies (DEST), IEEE*, pp. 1–6.

Kim, S.-H., Eom, J.-H., and Chung, T.-M. 2013. "Big Data Security Hardening Methodology Using Attributes Relationship," in *2013 International Conference on Information Science and Applications (ICISA)*, Ieee, June, pp. 1–2 (doi: 10.1109/ICISA.2013.6579427).

Jensen, M. 2013. "Challenges of Privacy Protection in Big Data Analytics," in *Proceeding of the International Congress on Big Data IEEE*, Ieee, June, pp. 235–238 (doi: 10.1109/BigData.Congress.2013.39).

Cardenas, A. A., Manadhata, P. K., and Rajan, S. P. 2013. "Big Data Analytics for Security," *IEEE Security & Privacy* (11:6), pp. 74–76 (doi: 10.1109/MSP.2013.138).

Marchal, S., Jiang, X., State, R., and Engel, T. 2014. "A Big Data Architecture for Large Scale Security Monitoring," in *Proceeding of the International Congress on Big Data IEEE*, Ieee, June, pp. 56–63 (doi: 10.1109/BigData.Congress.2014.18).

Xu, L., Jiang, C., Wang, J., Yuan, J., and Ren, Y. 2014."Information Security in Big Data: Privacy and Data Mining," *The Journal for rapid open access publishing* (2), pp. 1149–1176 (doi: 10.1109/ACCESS.2014.2362522).

Dev, H., Sen, T., Basak, M., and Ali, M. E. 2012. "An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks," in *Proceeding of High Performance Computing, Networking Storage and Analysis*, Ieee, November, pp. 1106–1115 (doi: 10.1109/SC.Companion.2012.133).

Wu, C., and Guo, Y. 2013. "Enhanced user data privacy with pay-by-data model," in *Proceeding of the International Conference of Big Data, IEEE*, Ieee, October, pp. 53–57 (doi: 10.1109/BigData

Gholami, A. and Laure, E., (2016) "Big Data Security and Privacy Issues in the Cloud." International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.1, January 2016 DOI: 10.5121/ijnsa.2016.8104 59

Liu, L., (2016)., "Security and Privacy Requirements Engineering Revisited in the Big Data Era" IEEE 55 DOI 10.1109/REW.2016.7

Gautam., and Rana, C. (2018) "**BIG DATA ANALYTICS: A SURVEY**" *International Journal of Latest Trends in Engineering and Technology Vol. (8) Issue (4-1), pp.288-293 DOI: http://dx.doi.org/10.21172/1.841.46 e-ISSN:2278-621X*

IDC, 2013. Big Data in 2020. In: IDC iView (Ed.):

IDC.

Patil, H.K., & Seshadri, R. (2014). Big Data Security and Privacy Issues in Healthcare. *2014 IEEE International Congress on Big Data*, 762-765.

Al-Barznji, K. and Atanassov, A. (2016), "**A MAPREDUCE SOLUTION FOR HANDLING LARGE DATA EFFICIENTLY**" **International journal for science, techniques and innovations for the industry MTM, YEAR X, Issue 12 / 2016, pp. 20– 23** ONLINE ISSN 1314-507X PRINT ISSN 1313-0226 http://stumejournals.com/mtm.html

Mill, K. A. (2017), "**What are the threats and potentials of big data for qualitative research?**" sagepub.co.uk/journalsPermissions.nav DOI:10.1177/1468794117743465 journals.sagepub.com/home/qrj

George, G. (2014) "**FROM THE EDITORS BIG DATA AND MANAGEMENT**" *Academy of Management Journal* 2014, Vol. 57, No. 2, 321–326. http://dx.doi.org/10.5465/amj.2014.4002

S. Lenka Venkata, "A Survey on Challenges and Advantages in Big Data," vol. 8491, pp. 115–119, 2015.

Sam B. Siewert. 2013. Big data in the cloud – Data velocity, volume, variety and veracity. https://www.ibm.com/developerworks/library/bd-bigdatacloud/index.html.

Fatima EL Jamiy, Abderrahmane Daif, Mohamed Azouazi, and Abdelaziz Marzak. 2015. The potential and challenges of big data-Recommendation systems next level application. CoRR abs/1501.03424 (2015). http://arxiv.org/abs/1501.03424.

A. Anton, E. Bertino, N. Li, T. Yu, A Roadmap for Comprehensive Online Privacy Management, Communications of ACM 50(7):109-116, July 2007.

Singh, J. (2014) "Real Time BIG Data Analytic: Security Concern and Challenges with Machine Learning Algorithm" 978-1-4799-3064-7/14/$31. 00©20 14 IEEE

Jayasingh, B. B., Patra M.R., and Mahesh D. B., (2016), "**SECURITY ISSUES AND CHALLENGES OF BIG DATA ANALYTICS AND VISUALIZATION**" *2016 2nd International Conference on Contemporary Computing and Informatics (ic3i)* 978-1-5090-5256-1/16/$31.00_c 2016 IEEE

A. Jacobs, (2009) The pathologies of big data, Communications of the ACM, 52(8), pp.36-44.

Damiani, E., (2015), "Toward Big Data Risk Analysis" 2015 IEEE International Conference on Big Data (Big Data) 978-1-4799-9926-2/15/$31.00 ©2015 IEEE

Bertino, E. (2015), "**Big Data – Security and Privacy**" 2015 IEEE International Congress on Big Data. DOI 10.1109/BigDataCongress.2015.126

Wamba, S. F., Aktar, S., Edwards, A., and Chopin, G. (2015), "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study" journal homepage: www.elsevier.com/locate/ijpe Int.J.ProductionEconomics165(2015)234–246

Rouse,M.,2011.Bigdata(BigData).Available from: ⟨http://searchcloudcomputing. techtarget.com/definition/big-data-Big-Data⟩ (retrieved11.03.13).

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U. (2015) "The rise of "big data" on cloud computing: Review and open research issues" Information Systems47(2015)98–115

C.Perera R. Ranjan, Lizhe Wang; S.U. Khan, A.Y. Zomaya, " Big Data Privacy in the Internet of Things Era", IT Pro May/June 2015

E. Bertino, R. Sandhu, Database Security – Concepts, Approaches, and Challeges, IEEE Transactions on Dependable and Security Computing 2(1): 2-19, January-March 2005.

Koutroumpis, P., & Leiponen, A. (2013). Understanding the value of (big) data. In ***Proceedings of 2013 IEEE international conference on big data.*** 38–42. Silicon Valley, CA, Los Alamitos, CA: IEEE Computer Society Press.

McKinsey Global Institute. (2011). ***Big data: The***

next fron-2014 *George, Haas, and Pentland* 325 ***tier for innovation, competition, and productivity.*** Lexington, KY: McKinsey & Company.

P. Institute, (2012). "Third Annual Benchmark Study on Patient Privacy and Data Security," Ponemon Institute LLC, 2012.

Lohr, S. (2012) "The Age Of Big Data" https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html accessed May 1st, 2019.

McKinsey Global Institute, (2011) 'Big Data: The Next Frontier for Innovation, Competition, and Productivity'.

Zheng, K., & Jiang, W. (2014). "A token authentication solution for hadoop based on kerberos pre-authentication" *In Data Science and Advanced Analytics (DSAA), 2014 International Conference on* (pp. 354-360).IEEE

Mills, K. A. (2017) "**What are the threats and potentials of big data for qualitative research?**" DOI:10.1177/1468794117743465

Polonetsky, J. and Tene, O., (2013), "PRIVACY AND BIG DATA: MAKING ENDS MEET" *STANFORD LAW REVIEW ONLINE* [Vol. 66:25