e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

SECURITY CHALLENGES OF BIG DATA COMPUTING

Mr. Gopala Krishna Sriram[*1]

[*1]Software Architect, Edge Soft Corp, Mckinney, TX USA.

ABSTRACT

In recent times, Big Data computing has become quite an important asset for firms. In fact, almost all industries are generating data in large amounts. Even though it has significant potential, it is still insecure and is a target of different security issues and problems. It has been identified in this research that Big Data computing requires great protection from different security challenges. There is a need for firms to improve their technological infrastructures, utilize advanced protocols, limit access, and even utilize a surveillance team for monitoring the status of data. It will enable firms to identify issues and resolve them quickly before confidential data can be accessed.

Keywords: Big Data Computing, Security Issues, Security Challenges, Technological Infrastructures, Advanced Protocols, Surveillance Team, Confidential Data, Etc.

I. INTRODUCTION

In almost every field, the most important asset for organizations has been 'big data' over the last few years. Data is the most important asset for not only computer science or technological industry but also other public as well as private organizations such as education, healthcare, or the engineering sector. To carry out the daily activities of companies, data is the essential thing, and businesses management can make the best decisions and achieve their goals on the information basis extracted from the data. It is estimated that 90 percent of the total data in recorded human history is created in the last recent years. Data's five exabytes were created by humans in 2003, and at present, this amount of information is created within two days.

Even though there has been a significant interest in big data and a large number of firms have adopted it, there are major security challenges associated with it. In fact, due to its security issues, businesses are concerned and they rely on the use of different techniques for improving security. In this paper, there will be a focus on the review of big data computing, its security concerns, and how these security concerns can be addressed [1].

II. RELATED WORK

Generally, Big Data is utilized for the management of datasets that are large in size and are beyond the ability of a common software to manage and analyze in an efficient manner. With the advancement in technologies, it is expected that the generation of data will double in the coming years [2]. The authors examined the strategic journey regarding big data privacy protection. The authors have stated that big data can be stored effectively and efficiently with the use of a number of strategies. However, the important thing is to consider the big data privacy lookup [3]. The authors for their research have surveyed the privacy techniques, obstacles, and requirements that are associated with the protection of data. An important role is being played by big data protection as it enables the confidentiality of data. Compared to data privacy, data security is different. Privacy concentrates only on a specific person using the data for making sure that it is being used in the right way. In Big Data analytics, privacy is essential due to a number of reasons. Weak protection techniques prove to be inefficient when it comes to Big Data [4].

Following environments might be capable of penetrating the privacy of a user in the technology of big data: First, during the transmission of data over the internet, personal information is shared with external sources. With this information, third parties might be able to harm the user. Secondly, personal data is sometimes collected for business purposes but it is exploited instead [5]. For instance, online shopping vendors seem to collect personal information, and using it, they can predict the habits and even activities of users. Fourthly, data trickling can take place in the processing and storing stages. In the third and second phases of the data lifecycle, data privacy is more significant. Typical database management systems are sufficient for the framework of big data because of the necessity of managing large volumes of data and data heterogeneity [3].

Privacy of big data faces many issues which are classified into 4 categories including Integrity Security, Data Administration, Data Privacy, and Data Security.

Framework security: The technology of Big Data follows the infrastructure of distributed computing and

several users work in parallel in it. It implies that the identification of intruders is very important. At present, most of the institutions have transferred to NoSQL databases from the traditional ones to handle semi-structured and unstructured data. NoSQL appears to offer architecture flexibility for the data that is multi-sourced but it is vulnerable to attacks.

Data Privacy: Various sources are used for collecting the data privacy has to be maintained in the analytic stage [6]. Encryption techniques can be utilized for protecting data.

Data administration: BD or big data is collected from countless sources making it contain numerous end-users. Gradually, complexity in big data increases. In big data, complexity will be concerned with provenance metadata because of the provenance graph.

Integrity security: Filtering process and input validation pose a significant challenge to the application of big data. Due to the data size, it is quite tough to determine whether the data is derived from a valid source or not. If the source is legit then the data has to be eliminated so that it doesn't possess a risk to the whole system. Security monitoring in real-time is chosen for alerting institutions at the primitive stages of attacks. SIEM systems also appear to play a significant role in helping the organization in identifying the issues and resolving them immediately [7].

In spite of the fact that information obtained through data mining can be quite useful for various applications, individuals have demonstrated an increasing concern about the coin's other side which is concerned with the threats of privacy posed by the technique. The privacy of an individual might be risked because of unauthorized access to personal information, the utilization of personal information for purposes not concerned with business, and undesired discovery of information that is embarrassing [8]. For example, Target received some complaints from an angry customer that coupons about baby clothes were sent to his teenage daughter. This information was actually obtained by mining the customer data. Analyzing this case, it can be analyzed that a conflict between privacy security and data mining is present. For dealing with the issues related to data mining, PPDM or privacy-preserving data mining has gained attention. Its objective is all about safeguarding personal information from unsanctioned or unsolicited disclosure while preserving the data utility [5].

PPDM's consideration is two-fold. First of all, sensitive data, like a cell number of a customer must be used directly for mining. Secondly, sensitive results of mining whose disclosure will be resulting a violation of privacy should be excluded. They examined information security in big data, including privacy and data mining [9]. Authors have shown their concern about the security of sensitive information of individuals threatened by the development and growing popularity of data mining technologies. Authors have identified the four distinct kinds of users involved in the application of data mining: (i) data provider, (ii) data collector, (iii) data miner, and (iv) decision-maker. Authors have reviewed the game-theoretical approaches as well as explored the approaches regarding privacy-preserving for every kind of user. Authors have provided useful insights into the PPDM's study by differentiating the different users' responsibilities concerning the sensitive information's security [10].

The authors critically analyzed the big data challenges, along with analytical methods. A holistic view of big data practices is presented by the authors and the big data applications in the normative literature slice [11]. Authors have adopted the SLR methodology that is, according to the authors, a most convenient tool to conduct the descriptive review of existing literature. The authors have briefly explained the challenges in big data through synthesizing and systematic analysis of literature. The authors have briefly explained the data, management, and process challenges regarding big data. All the procedure was done to provide a useful direction to future research [6].

Although the advantages of big data are both substantial and factual, there still are numerous issues and challenges that have to be addressed for fully realizing the actual potential of big data [12]. Addressing them will be quite helpful in realizing the true capability of big data and utilizing it to its maximum potential. It can be said that some of the issues are a function of the specifications of big data, some, bits present analysis models and analysis, and some, through the boundaries of the present system of data processing [13]. Therefore, there are different aspects that are the reason for prevalent issues. Extant studies which surround the challenges of big data have actually paid attention to all the difficulties of recognizing the notion of big data, decision making of the information which is collected and generated, problems related to privacy, and all ethical considerations

which are associated with the mining of such data [14]. It is actually asserted by authors that creating a sustainable solution for multifaceted and large data is quite an issue that businesses are facing and trying to resolve it by consistently learning and then applying innovative and new approaches. For instance, one of the largest issues about the infrastructure of big data is high costs. Equipment of hardware is quite costly even with the presence of technologies of cloud computing [12].

In addition, for sorting through the data, so that important data can be developed, human analysis is required frequently [1]. Although the technologies of computing are needed for facilitating these data, keeping up the pace, talents, and human expertise required by business leaders for leveraging big data are still lagging behind, which proves to be another significant issue. Just as the authors suggest, some issues of big data can be classified into three categories on the basis of the lifecycle of data: Data challenges relating to the specifications of data such as dogmatism, discovery, quality, veracity, velocity, variety, and volume etc. Challenges of the process are associated with how techniques: how results can be provided, how can the right analysis model be selected, how data can be transformed, how data can be integrated, and how it can be captured. Challenges of management cover, for instance, ethical, governance, security, and privacy aspects [14].

In the context of Big Data computing, a major challenge is concerned with the management of information while handling rapid and massive data streams [15]. Therefore, there is a need for security tools to be scalable and flexible for simplifying the incorporation of technological evolutions and managing the changes that might occur in the requirements of applications. In addition to it, there is a need for finding a balance between dynamic analysis, system performance, and multiple requirements of security. It should be noted that traditional techniques of security like data encryption tend to decrease performance and they also consume a significant amount of time. At the same time, these techniques are not efficient. Therefore, most of the time, the attacks on security are identified after the damage has been done and sustained [16].

The platforms of Big Data tend to imply and indicate the management of parallel computations and various applications. Thus, for real-time analysis and data sharing, the key element is performance. The combination of different techniques and methods might bring hidden risks and issues that are mostly underestimated and not evaluated [17]. Therefore, the platforms of Big Data might bring new security vulnerabilities and risks that are not evaluated. In addition to it, the value of data is concentrated on a number of data centers and clusters. These rich data mines are quite attractive for industries, governments, and commerce. There is no doubt that they constitute a target of several penetrations and attacks [18]. At the same time, most of the security risks tend to come from end-point users, partners, and employees. Therefore, there is a critical need for the deployment of advanced mechanisms for the protection of clusters of Big Data. In this regard, there is a responsibility of data owners to set clear policies and clauses associated with security [19].

For ensuring data security and privacy, there is a need for achieving data anonymization without influencing the quality of data or the performance of the system. Traditional techniques of anonymization, however, are based on a number of computations and iterations that consume significant time. In addition, several iterations tend to influence data consistency and they also decrease the performance of the system, especially when heterogeneous data sets are to be managed [20]. It is quite difficult to analyze and process Big Data when it is anonymized. It is worth noting that some security techniques and methods are not compatible with technologies of Big Data such as the MapRecude paradigm [21]. For ensuring the security of Big Data, there is a need for verifying the compatibility of different security technologies with Big Data methods. Actually, the reliability and precision of data analysis tend to depend on the integrity and quality of data. Thus, there is a need for verifying the integrity and authenticity of sources of Big Data before the data is to be analyzed. Considering the fact that large volumes of data are created on a consistent basis, it is quite tough and complex to assess the integrity and authenticity of all the data sources [22].

In addition to it, for extracting complete information from different sources of Big Data, there is a need for analysts to manage heterogeneous and incomplete data streams that exist in different formats. They are required to filter data in an efficient manner and they have to contextualize and organize data as well before they can perform an evaluation of the data. Government agencies and private organizations have to respect a number of industry standards and security laws that have the objective of enhancing and improving the management of security and confidentiality of data. It is, however, important to note that some ICTs might even

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:01/January-2022 Impact Factor- 6.752 www.irjmets.com

involve different entities across a number of nations. Therefore, enterprises have to manage different regulations and laws as they operate. Big Data analytics might conflict with a number of privacy guidelines and concepts. For instance, different data sets can be correlated by analysts from different entities for revealing sensible or personal data with the use of anonymization techniques. As a consequence, such analyses might help in identifying confidential information [23].

In the case of social networks, there is no doubt that a huge amount of comments, videos, photos, and clicks are created with the use of social networks. Usually, they are the first or primary source of information for a number of entities. On social networks, Big Data constitute an important mine for different governments to better analyze and manage national security issues and risks. In fact, some governments and associations tend to evaluate social networks for supervising public opinions. Still, the analysis of such data is not simple. It requires computation power that is not really possessed by an individual firm. For the enhancement of security of Big Data, firms and organizations tend to depend on advanced analysis of dynamic security [24]. The underlying objective is concerned with analyzing and extracting security events in real-time for enhancing transactional and online security for the prevention of attacks. It is important to note that some of the common techniques used by firms for the protection of Big Data include:

Anonymization: A common technique that is used by firms and organizations for the protection of data is data anonymization. It helps in protecting data across distributed and cloud systems. A number of solutions and models are utilized for the implementation of this technique including l-diversity, k-anonymity, m-invariance, and t-closeness. The sub-techniques are based on Bottom-Up Generalization and Top-Down Specialization [7].

Data Cryptography: Another common technique that is used for the protection of data is data encryption. It is utilized for ensuring the confidentiality of Big Data. In contrast with typical techniques for encryption, it should be noted that Homographic Cryptography allows computation even on the data that is encrypted. As a consequence, this method ensures the confidentiality of information which enables the extraction of insights through computations and analyses [25].

Centralized Security Management: A number of firms make the use of the cloud for the storage of data. The underlying goal is concerned with taking advantage of a centralized security mechanism and the standard compliance infrastructure [26]. Still, it is quite difficult to achieve the status of zero risk. The cloud tends to get the attention of hackers and attackers because it represents a base or hub of confidential information [10].

Data Access Monitoring

Actually, there is an increasing rise of security issues and threats due to the increasing rate of exchange in data over the cloud and distributed systems. For facing these challenges, it has been proposed that controls at the data phase should be integrated [27]. However, it has also been identified that this integration is not sufficient for combatting security threats. In addition to it, there is a need for access controls to be granulated in such a manner that access is limited by responsibilities and roles. There generally exist a number of methods for ensuring data confidentiality and access control including federated identity management, smart cards, and certificates. The consistent monitoring of security threats can prove to be efficient in the sense that threats and problems can be identified quickly and they can be managed without experiencing major difficulties and issues [28].

Security Surveillance

There is no doubt that there exists a need for ensuring consistent surveillance or detecting security incidents and related issues and problems in real-time. For ensuring the surveillance of Big Data security, there are a number of solutions that can be considered and used by firms. These solutions include dynamic analysis of security threats, Security Information and Event Management, and Data Loss Prevention. These solutions are actually based on correlation and consolidation methods between different data sources. There is also a significant need to carry out regular audits for performing and verifying different security policies and the recommended practices for employees and users [29].

III. METHODOLOGY

Generally, for every research, there are a number of methods and techniques that are considered and utilized for performing research. In this case as well, there are some certain methods that have been considered and

utilized. These methods are primarily concerned with qualitative methods. Actually, qualitative and quantitative methods or techniques are recognized as two primary methods of research that are available for researchers to utilize. In fact, it would not be wrong to say that these methods are often considered by authors for performing their research and conducting their studies. Both of these methods and techniques are quite different from each other. In this case, qualitative methods have been considered in the form of literature reviews.

Literature reviews are recognized as a common form or type of qualitative technique and it can be said that they are mostly considered due to their efficiency and their simplicity. Even though they are quite simple, they deliver the outcomes that are required in an efficient manner. In fact, it would not be wrong to say that literature reviews enable a researcher to ensure that the topic or concept at hand is explored in an effective manner. For instance, it presents and provides a researcher with an authentic and systematic method of exploring different researches and studies. Through this thorough exploration, the concept or topic can be detailed in an effective manner. In this case, a literature review has been selected and performed because it seems to suit the nature of the research.

Actually, quantitative methods in the form of questionnaires could have been considered if the timescale of the research had been extensive. In addition to it, it would have been considered if the nature of the research had been different. However, in this case, the use of questionnaires for performing surveys was not considered suitable. In fact, it can be said that their use was not considered to be effective. It would not have delivered the required outcomes. In fact, it can be said that if questionnaires had been considered, it would have caused and resulted in quantitative data. However, it would not have resulted in qualitative and conceptual information. Therefore, in this case, quantitative methods have not been considered.

Typically, such methods and techniques are considered and utilized when there is a need to ensure that quantitative information is obtained on the topic at hand. In this research, there was not actually a need to obtain quantitative information. Instead of it, there was a need to evaluate the concept and in order to do it, literature reviews were recognized and considered to be the most suitable technique of performing the research. In fact, it would not be wrong to say that the use of literature reviews enabled the exploration of the topic in an effective manner. A number of credible journal articles were considered and they were thoroughly examined for performing the research and obtaining all the necessary information that was required to be included in this study [28].

IV. DISCUSSION

In accordance with the information obtained from the literature review, it can easily be said that big data has a number of benefits for organizations. In fact, it enables businesses to perform operations and tasks that were not really possible before. However, similar to how different benefits are offered by big data, there are also some drawbacks, and one of the key drawbacks of big data is actually associated with its weak privacy. There are undoubtedly some significant privacy issues and risks that are present when it comes to big data. Actually, in modern times where processes such as personalized marketing and filter bubble are present, many people fear that their privacy is at risk [16].

It is important to note that a large part of insights associated with big data includes predictions that are made regarding the details of consumers. In fact, most of the time, these details are quite personal in terms of their nature, which is one of the reasons why even the chance or possibility of them falling into the hands of wrong people, is enough to eliminate any possible trust that consumers have in different firms and organizations. Realizing the importance of privacy and the value that is possessed by sensitive information, it is necessary to the survival of a firm that they consider different measures for preventing the obstruction of privacy of consumers [4].

In addition to it, in spite of obtaining and receiving significant scrutiny, thorough and complete anonymity over the web is more than just a little difficult. Big data analytics are considered by firms and due to it, the possibility of using anonymous files becomes impossible. Considering the fact that big data insights are generally based on different types of raw datasets, there exists a significant possibility that consumers might have their identity exposed. Even if there is a chance that a data file is completely anonymous, a number of security teams seem to consider and combine these data files to ensure that connections can be made. Thus, the identification of a

person is made quite simple and easy. Furthermore, being anonymous is complicated more by the fact that almost every SME that performs online business depends on the software that is hosted by different third parties in the cloud. These firms have different privacy practices which again presents a significant risk to privacy [8].

In a possible attempt to secure and protect their sensitive information from cybercriminals and hackers, most firms make the use of data masking as a procedure. It is actually a process through which actual information is hidden by less important data sets and information. Normally, data masking is considered and utilized for veiling sensitive information different unauthorized people. In most firms, the primary or basic function that is served through data masking is actually the protection or security of confidential information from being exposed and leaked. If it is not considered and used properly, data masking is capable of compromising security to a significant extent [29].

Actually, the use of big data is increasing significantly and it would not be wrong to say that firms make the use of big data for a number of purposes. However, big data experience different privacy concerns and risks. As it has been discussed above, the protection of big data is not simple and it is quite a complex process. For the protection of big data, it is important to note that there are a number of steps that are required to be considered and taken. One of the most important steps for a firm to consider is to have the necessary infrastructure for the management of big data. Without the required infrastructure, it is highly possible that the intended data might be exposed to the wrong people who could make the use of this information for their own specific purposes [23].

Other than having the required infrastructure, it is necessary for firms to make sure that the right security practices are in place that can help in securing big data. With the presence and use of different security protocols, big data can be protected. However, it increases the cost of utilization of big data to a significant extent. In such a case, firms have to invest a significant amount of their revenues and money into the security and protection of big data.

V. CONCLUSION

Overall, it can be said that while big data offers some significant benefits to firms, it also poses some significant issues and challenges. One of these challenges is concerned with the sheer number of privacy risks that are experienced by it. The use of big data is quite risky in the sense that the identity of consumers can be exposed, and it can eliminate all the possible trust they have in a specific brand or a firm. For the protection of big data, there are generally a number of steps that are required to be taken by firms. One of these steps is concerned with the utilization of effective infrastructure that is capable of managing big data without any type of issue. The second step is actually associated with the use of different security protocols that are necessary for securing big data and ensuring that big data is protected in an effective manner.

Third, there is a need of utilizing techniques that can be scaled according to the requirements and needs of the firms. Without this scalability, it would not be possible to protect all the areas within the firm. Fourth, there is a critical need to ensure that there is a surveillance team within the firm that is responsible for monitoring the status of security within the firm. It will enable firms to identify and evaluate security threats and issues that tend to exist in the organization. In addition, it will enable the team to analyze the attacks and threats quickly before performing and implementing the strategies that can help in ensuring that the threats are addressed in an efficient manner without experiencing any difficulties. At the same time, there is a need for enhancing the security protocols within the firms consistently.

VI. REFERENCES

[1] J. Moreno and M. A. Serrano, “Main Issues in Big Data Security,” Future Internet, vol. 8, no. 44, pp. 1-16, September 2016.

[2] IDC, December 2012. [Online]. Available: https://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm.

[3] D. Chen and H. Zhao, “Data security and privacy protection issues in cloud computing,” 2012 International Conference on Computer Science and Electronics Engineering, pp. 647-651, 2012.

[4] M. P. Babu and S. H. Sastry, "Big data and predictive analytics in ERP systems for automating decision making process," 2014 IEEE 5th International Conference on Software Engineering and Service Science, pp. 259-262, 2014.

[5] W. El-Hajj, "The most recent SSL security attacks: origins, implementation, evaluation, and suggested countermeasures," Security and Communication Networks, vol. 5, no. 1, pp. 113-124, 2012.

[6] F. L. Greitzer, A. P. Moore, D. M. Cappelli, D. H. Andrews, L. A. Carroll and T. D. Hull, "Combating the insider cyber threat," IEEE Security & Privacy, vol. 6, no. 1, pp. 61-64, 2008.

[7] M. D. Viji, K. Saravanan and D. Hemavathi, "A Journey on Privacy protection strategies in big data," IEEE, pp. 1344-1347, 2017.

[8] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins and N. Kruschwitz, "Big data, analytics and the path from insights to value," MIT sloan management review, vol. 52, no. 2, pp. 21-32, 2011.

[9] C.-C. Lee, P.-S. Chung and M.-S. Hwang, "A Survey on Attribute-based Encryption Schemes of Access Control in Cloud Environments," IJ Network Security, vol. 15, no. 4, pp. 231-240, 2013.

[10] L. Xu, C. Jiang, J. Wang, J. Yuan and Y. Ren, "Information Security in Big Data:Privacy and Data Mining," IEEE, vol. 2, pp. 1149-1176, October 2014.

[11] L. Leichtnam, E. Totel, N. Prigent and L. Mé, "Starlord: Linked security data exploration in a 3d graph," 2017 IEEE Symposium on Visualization for Cyber Security (VizSec), pp. 1-4, 2017.

[12] J.-h. Li, "Cyber security meets artificial intelligence: a survey," Frontiers of Information Technology & Electronic Engineering, vol. 19, no. 12, pp. 1462-1474, 2018.

[13] G. Martin, P. Martin, C. Hankin, A. Darzi and J. Kinross, "Cybersecurity and healthcare: how safe are we?," Bmj, 2017.

[14] R. Mittu and W. F. Lawless, "Human factors in cybersecurity and the role for AI," 2015 AAAI Spring Symposium Series, 2015.

[15] J. Norbekov, "Ensuring information security as an ideological problem," Mental Enlightenment Scientific-Methodological Journal, pp. 56-65, 2020.

[16] J. C. Ogbonna, F. O. Nwokoma and A. Ejem, "Database Security Issues: A Review," International Journal of Science and Research, vol. 6, no. 8, 2015.

[17] M. V. Pawar and J. Anuradha, "Network security and types of attacks in network," Procedia Computer Science, vol. 48, pp. 503-506, 2015.

[18] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya and Q. Wu, "A survey of game theory as applied to network security," 2010 43rd Hawaii International Conference on System Sciences, pp. 1-10, 2010.

[19] C. Raleigh and C. Dowd, "Governance and conflict in the Sahel's 'ungoverned space'," Stability: International Journal of Security and Development, vol. 2, no. 2, 2013.

[20] P. Russom, "Big data analytics," TDWI best practices report, fourth quarter, vol. 19, no. 4, pp. 1-34, 2011.

[21] P. W. Singer and A. Friedman, Cybersecurity: What everyone needs to know, OUP USA, 2014.

[22] M. Siponen, M. A. Mahmood and S. Pahnila, "Employees' adherence to information security policies: An exploratory field study," Information & management, vol. 51, no. 2, pp. 217-224, 2014.

[23] U. Sivarajah, M. M. Kamal, Z. Irani and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," Journal of Business Research, vol. 70, pp. 263-286, 2017.

[24] Z. A. Soomro, M. H. Shah and J. Ahmed, "Information security management needs more holistic approach: A literature review," International Journal of Information Management, vol. 36, no. 2, pp. 215-225, 2016.

[25] A. Strielkina, O. Illiashenko, M. Zhydenko and D. Uzun, "Cybersecurity of healthcare IoT-based systems: Regulation and case-oriented assessment," 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), pp. 67-73, 2018.

[26] H. Xiao, B. Ford and J. Feigenbaum, "Structural cloud audits that protect private information," Proceedings of the 2013 ACM workshop on Cloud computing security workshop, pp. 101-112, 2013.

[27] D. Zhe, W. Qinghong, S. Naizheng and Z. Yuhan, "Study on data security policy based on cloud storage," 2017 ieee 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids), pp. 145-149, 2017.

[28] N. Walliman, Research methods: The basics, Routledge, 2017.

[29] P. Zikopoulos and C. Eaton, Understanding big data: Analytics for enterprise class hadoop and streaming data, McGraw-Hill Osborne Media, 2011.