Analysis of Vulnerabilities in Hadoop Map Reduce Framework: A Review

Shubham Jambhulkar [a], Deepak Singh Tomar [a] and R K Pateriya [a]

[a] *Maulana Azad National Institute of Technology, Bhopal, India*

Abstract

Enormous Data is an assortment of various equipment and programming advancements, which have a heterogeneous framework. Hadoop system assumes the main part in managing and putting it away. It provides intelligent financial and fast data applied in various regions such as clinical benefits, social networks, and safeguard. Hadoop Framework is based on distributed streaming model and is used to manage and store data within wide range of product PCs. Because of the adaptability of the system, a few weaknesses emerge. These weaknesses are dangers to the information and lead to assaults. In this paper, various sorts of weaknesses are talked about and potential arrangements are given to diminish or take out these weaknesses. The test arrangement used to perform normal assaults to comprehend the idea and execution of an answer for staying away from those assaults is introduced. The outcomes show the impact of assaults on the presentation. As per results, there is a need to ensure information utilizing guards inside and out to security.

Keywords

Big Data, Map-Reduce, Hadoop, Vulnerability, Kerberos

1. Introduction

Big Data is a gathering of exceptionally enormous informational collections [1] which are extremely perplexing or too huge to ever be worried about by customary information handling applications. For any data to be regarded as big data it must satisfy the 4 V's namely Velocity, Veracity, Volume and Variety [3], [2]. With the advancement of technology in today's world, a large amount of information is produced in various fields such as social networking sites, transaction records, data sensors, log files, etc. Due to this various source terabytes of assembled, semi-assembled, and unassembled data are produced at every point of time. Therefore, if this data is not stored or pre-processed there is a chance of loss of this important data. To avoid this loss, the Hadoop framework is used with different analytics tools and they are often much quicker than conventional analytical methods of the past.

Big data is a word that is equally associated with Hadoop. As previously discussed, for any data to be considered as big data it must satisfy 4 V's namely Velocity, Veracity, Volume and Variety. With data advancement, it had not only affected the Velocity, Veracity, Volume and Variety aside from the privacy and security aspects in data. Additionally, the inclusion of another V that is Vulnerability is proposed [4]. Figure 1 represent the different V's in a diagrammatic manner.

WCNC-2022: Workshop on Computer Networks and Communications, April 22 – 24, 2022, Chennai, India.
EMAIL: shubham.pj0806@gmail.com (Shubham Jambhulkar)
ORCID: 0000-0002-2934-3934 (Shubham Jambhulkar)

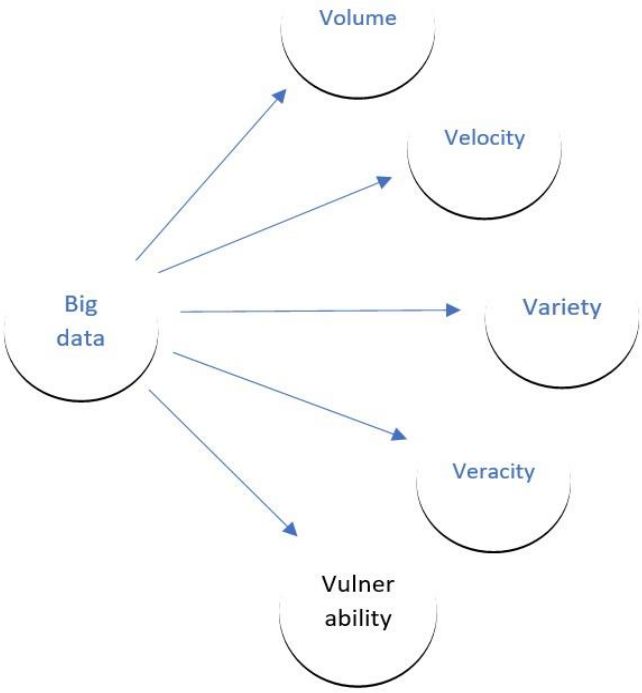

Figure 1: Big data representation of different V's [2], [3]

Big data accumulates all the interesting values from the data pool and many countries are operating on dominant schemes on the basis of big data. As a result of global level schemes many latest models and framework have been developed. Some frameworks were developed for providing a considerable and good amount of storage capacity, real-time data analysis, and parallel processing of data [5]. One such popular framework is Hadoop. The advantages offered by big data are very vast. The technology offers better scalability, flexibility, with fulfillment-based in a affordable rates. Subsequently recent growth in sustainable technology, the cost associated with the processing and storage section continues to decrease [6].

The recently developed technology is designed to guarantees privacy and security aspects in comparison to traditional previous technologies. But even with these advantages, they are becoming prone to negative purposes. With the recent growth in fields and organizations using this technology for storing and processing their private organization's data, it has become prone to negative data attacks.

1.1. Hadoop Framework

Apache Hadoop provides a way to process parallelly the same distribution of very large or complex databases. The Hadoop framework provides advantages like distributed computing and parallel processing for datasets. Hadoop comprises a component such as HDFS, Map Reduce, and YARN. HDFS supervises the repository, Map Reduce supervises processing in parallelly and YARN is responsible for resource management in Hadoop's Cluster.

1.1.1. Hadoop

During 2005 Hadoop initially appeared and was introduced at later 2011 to help spread the web searching tool scheme to Yahoo. [7]. The delivery had very little safety assist, made for people who were loyal to the Climate. Hadoop since then has emerged among the modern state-of-the-art advancements to store, process, and examine large information through utilizing bunch of out-spread

climate [8]. The Framework clients subsequently unrolled from one side of the planet to the other, generally enormous organizations [9].

1.1.2. Hadoop Distributed File System

Hadoop distributed file system is responsible for the storage mechanism of the Hadoop framework and can operate without any hurdles. In HDFS, a big complex file is distributed over the cluster network that is comprised with multiple nodes of data and associated repository. During this cycle segments of the Node Name, the first record becomes square, 64 MB in size and repeats on various Data Nodes depending on the rules with the previous characters. Name Node additionally comes with metadata for this duplication and distribution. Every information block has been redesigned multiple times for maximum access, two from one Data Node site and one from various Data Node racks. The group Information Node stores a small portion of all text. Name Node always remembers which information block has the location where the file is located, where the information blocks are set, and where the power limits are involved. Using periodic signals, Name Node invariably knows which Data Nodes are still available. When the signal (heartbeat) is missing, the Name Node detects a Data Node failure, eliminates the Data Node bombed in the Hadoop group, and attempts to distribute the information load evenly across the current Data Node. Alternatively, the Name Node ensures that a specified number of duplicates of information is kept constant for maximum access. The diagram below (Figure 2) shows the Hadoop distribute file structure architecture

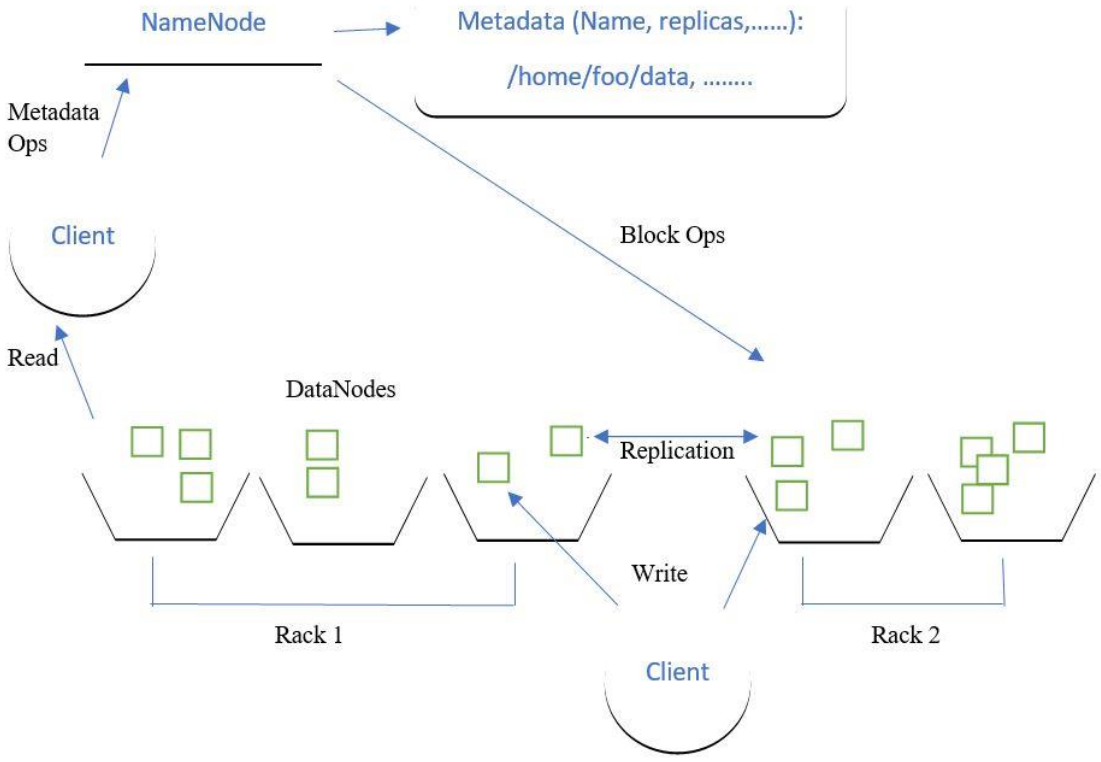

Figure 2: HDFS Architecture

1.1.3. Map Reduce

MapReduce is an equal handling structure work dependent on the expert slave guideline, like Hadoop Distributed File System. Map Reduce is mixture consists of three slave agents per slave and one expert agent in group. Map Reduce management depends equally on various calculations for direction and downtime. This works in two stages, the map function and the reduction function. This JobTracker divides the database into separate clusters called map operations and directs them into

three Data Nodes naturally across all related product computers distributed across the organization for equal management.

Let us have a look at block diagram of Map reduce phases in Figure 3.

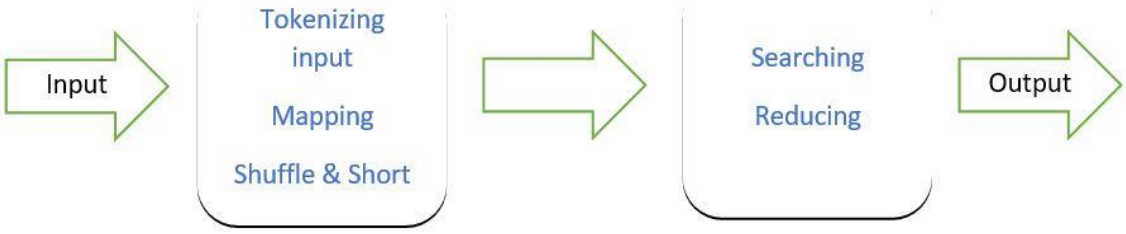

Figure 3: Phases in Map Reduce namely Shuffling and Reducing

Ordinarily, the guide assignments run on a similar bunch of Data Nodes where information lives (Data area). Assuming a hub is as of now vigorously stacked, another hub that is near the information, i.e., ideally a hub in a similar rack, is chosen. Moderate outcomes are inaccessible to the client and are traded among the nodes (Shuffling), and from there on, converged by the decreased undertakings to get the outcome. The Figure 4 shown below shows the internal algorithm of Map Reduce.

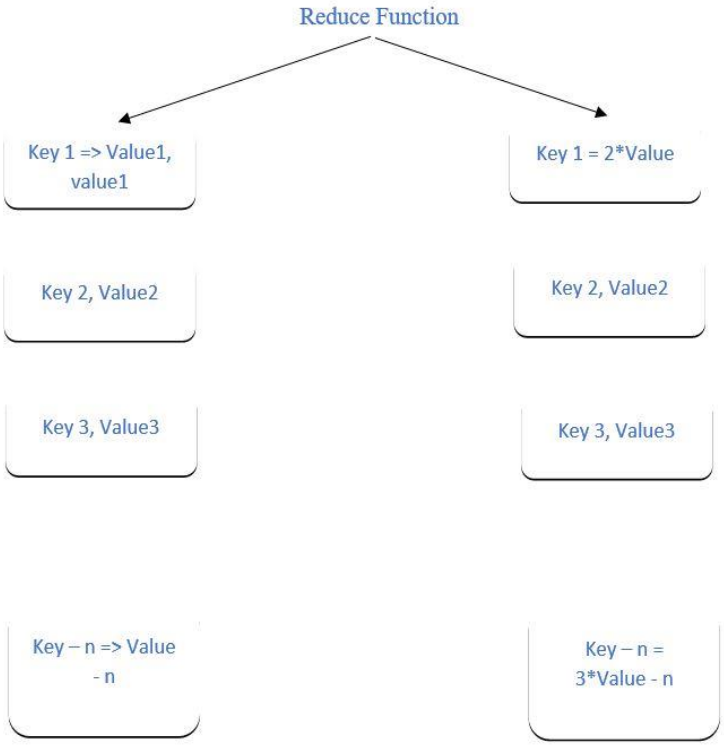

Figure 4: Algorithm for Mad Reduce Function

The Table 1. Shows the internal structure Key value pair in the Map Reduce.

Table 1: Map Reduce phases Structure

Phase	Input	Output
Mapper	(Key, Value)	(Key, Value)
Shuffle & Sort	(Key, Value)	(Key, list (Value))
Reducer	(Key, list (Value))	(Key, Value)

Transitional aftereffects of guide stages were amassed with storing short information size as conceivable within the transfer undertakings to diminish assignments. Medium results are stored in the nearest Data Node record system. JobTracker responds by carefully resetting any function in the event of a disruption. If an undertaking doesn't advise any advancement is still up in the air time, or on the other hand, assuming Node of data flops totally, whole assignment will be booted on other server counting errands however, will not be wrapped up. Assuming an errand runs very leisurely, the JobTracker likewise restarts the assignment on one more server to execute the general occupation at a suitable period.

1.1.4. Yet Another Resource Negotiator

MapReduce was split into two categories: Yet Another Resource Negotiator and Map Reduce [7]. Yet another Resource Negotiator primary rule is to isolate the assets of the executives and occupation planning functionalities into independent daemons. An asset administrator refers assets among framework implementations, with hub chief assistance. The asset director has two fundamental parts: application supervisor and scheduler. scheduler assigns assets to different operating systems and books depending on the asset requirements of the applications. The asset director acknowledges work entries, and each occupation is distributed to the application supervisor. The diagram below figure 5 represents the YARN log file architecture.

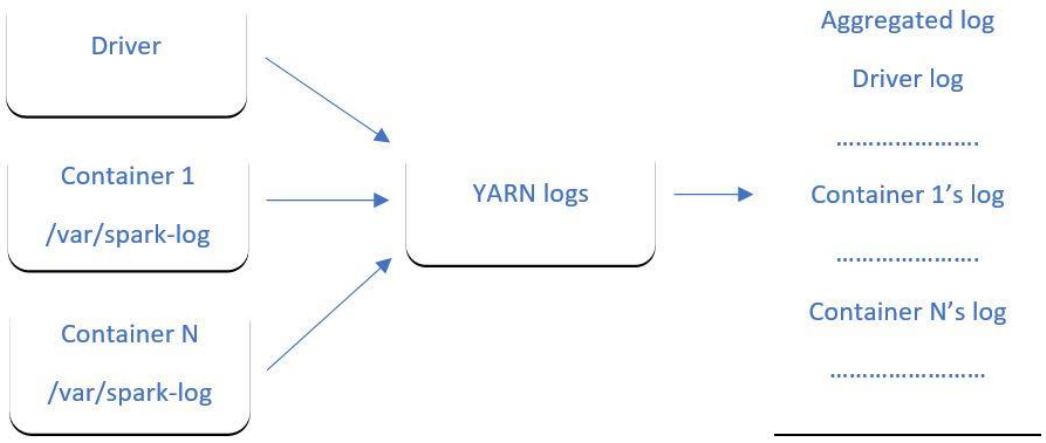

Figure 5: YARN log file Architecture

2. Related Work

This section reviews and provide an analysis of different vulnerabilities on the Hadoop framework. Generally, our concerned area lies in the Map-Reduce functionalities provided by the Apache Hadoop framework. There has been numerous tools and method defined to tackle the problem of vulnerabilities but every tools is able to provide one functionality while leaving several other holes in the system. The use of Kerberos system with proper authorization is suggested to tackle multiple problems responsible for vulnerabilities.

3. Literature Review

In a report [10], the vulnerability is categorized as data privacy, infrastructure security, and data management. These classifications are further classified into three different categories: Dimensional modeling, architecture dimension, and information flow. The life cycle of data comprises data in transit and data privacy comprises data at rest.

As per another report [11], the big data security and privacy issues are categorized into five types namely, Hadoop Security, Key management, anonymization, monitoring, and auditing. The author also proposed some algorithms concerning security and monitoring aspects of sensitive information like the Bull eye algorithm.

As per another report [12] on cloud security, a security model with some proposed cloud infrastructure layered was designed. This model was then further classified into four categories: logical, basic, governance, and value-added security. This report specifies the infrastructure policy framework of Hadoop.

As per another report [13], the author specifies different types of attacks that have taken place in the Hadoop framework. The attacks namely comprised of Denial of Service, Man in the Middle, impersonation, repudiation and replay attacks. According to the author, because of the distributed nature of the Map-Reduce component of Hadoop possible wide range of attacks were possible leading it to a vulnerable state. The ideal Map Reduce component would be comprised of proper authentication control, access control, authorization, confidentiality of data, and lastly data availability for Map and reducer class of Map-reduce. For better authentication control the author recommends the use of Kerberos protocol.

From one more report [14], the security and privacy aspects faced challenges that were categorized in different model names namely access control, access control policy, Data confidentiality, and lastly smart objects. This report puts forwards the challenges of research faced in regards to comprehensive solutions for securing security and privacy aspects.

Another report [15] lists out the challenges faced when the privacy and security aspects are needed to be ensured to be safe. The challenges were broadly classified into Risks concerning privacy, Credibility of data, lacking of recent technologies, and threats. To cope with these challenges author introduces supervising data, protection mechanism, protection agency, and quality of data.

As per another report [16], different categories of security and privacy aspects and the connection between them were discussed briefly. The aspects were classified as Confidentiality, analytics, integrity, privacy, stream processing, data format and lastly visualization.

This report [17], has showcased an investigation with the corporate perspectives relying on big data aspects simply and most effectively. Accordingly based on this corporate perspectives economic perspective, investment decisions, fighting cybercrimes, and cyber insurance. The Table 2 represents the vulnerabilities reported in online databases as shown below.

Table 2: Tabular representation of attacks described in online database [18].

Year	Total Vulnerabilities	Denial of Service	Cross-Site Scripting
2011	44	15	7
2012	63	19	6
2013	74	25	9
2014	92	23	6
2015	57	19	5
2016	103	15	17
2017	217	29	22
2018	148	15	9
2019	158	13	14
2020	161	6	16
2021	193	16	10

3.1. Vulnerability Databases

There are numerous online databases currently available all over the internet, that are mainly responsible for exposing the possible security vulnerabilities on numerous products and hardware. There are numerous such online databases namely, Common Vulnerabilities and exposures, Computer emergency readiness team, National Vulnerability databases, and Open-Source Vulnerability databases.

The CVEs uniquely identify the vulnerabilities based on an identification number. Based on CVE the list of vulnerabilities encountered in Hadoop has been shown in the tabular format below [19]. The Table 3. shows different vulnerability reported in CVE.

Table 3: Detailed CVEID of various attacks published on online database[4], [19].

CVE ID	Description
CVE-2021-45911	An issue was discovered in gif2apng 1.9. There is a heap-based buffer overflow in the main function. It allows an attacker to write 2 bytes outside the boundaries of the buffer.
CVE-2021-45906	OpenWrt 21.02.1 allows XSS via the NAT Rules Name screen.
CVE-2017-7669	In Apache Hadoop 2.8.0 the LinuxContainerExecutor runs docker commands as root with insufficient input validation. When the docker feature is enabled, authenticated users can run commands as root.
CVE-2017-3162	HDFS clients interact with a servlet on the Data Node to browse the HDFS namespace. The Name Node is provided as a query parameter that is not validated in Apache Hadoop before 2.7.0.
CVE-2017-3161	The HDFS web UI in Apache Hadoop before 2.7.0 is vulnerable to a cross-site scripting (XSS) attack through an unescaped query parameter.
CVE-2017-15713	Vulnerability in Apache Hadoop 3.0.0 allows a cluster user to expose private files owned by the user running the MapReduce job history server process. The malicious user can construct a configuration file containing XML directives that reference sensitive files on the MapReduce job history server host.

3.2. Patch Management

Patch management is a mechanism for detecting and eliminating the vulnerabilities before any attackers try to exploit them. The throughput is directly proportional to the fast detection of vulnerabilities, rectified, compressed with some methods like scanning and testing for reviewing of code. As per the report from 2017, the application of scanning methods has been gradually rising internationally [20].

3.3. Security Issues in Hadoop

It is known that Hadoop was designed primarily for a performance basis and not on security basis. The Developers decided that security functionalities will be added over time to increase the framework efficiency. Due to this the security mechanism of Hadoop was very weak and prone to many attacks. Hadoop was mainly designed with a focus on improving efficiency. But due to recent

attacks researchers are now focusing on the security aspects of Hadoop. However, presently there does not exist any evaluation method for the security policies of Hadoop.

Due to the recent growth of Big Data, the security policies available are not up to the benchmark to be even considered for evaluation. The ecosystem of Hadoop comprises a collection of different applications, where every application requires some security mechanism to function accordingly for Big Data.

Out of all the models proposed previously to work with big data, Hadoop was uniquely identified because of its distribution system with parallel processing but was lacking in the security policies. Whereas, the distributed nature of Hadoop was favored previously, now the distributed nature of computing is posing a set of new vulnerabilities for professional and security managers [21].

3.4. Security threats and possible attacks

Any possible danger for the information system can be referred to as a threat. A threat is basically what an attacker tries to identify and use as an attack against any company or organization [22]. Also are already familiar with the CIA triad. For any system to be regarded as secure it must satisfy Confidentiality, integrity, and Availability also known as CIA triads. To comply with confidentiality, an authenticated server can be implemented that can access the whole system.

3.4.1. Impersonation Attacks

This type of attack occurs when an attacker tries to impersonate the registered or legitimate authority for accessing the resources. The attackers can make use of different sets of tools and methods to steal sensitive information attacks directly on the Hadoop Clusters leaving the system vulnerable. To perform an impersonation attacks an attacker can try to replay the acknowledgement received from Kerberos protocol. At last, when the attackers gain access to the Hadoop framework, performing actions like leaking and throttling the processing time of Map Reduce.

3.4.2. Denial-of-Service Attacks

A Denial-of-Service [23] is a type of attack where an attacker floods the system with an enormous request which makes the system unable to allocate resources to legitimate users. As per the report, more than 11247 attacks have been taken place among which 5 attacks were able to breach the security. Denial of Service attacks is basically where a system is flooded with large request or traffic causing the system servers to crash or halt all the operations. Denial of Services can be initiated in two ways: by crashing the services, flooding the services. The Hadoop Component like Name Node and the authentication server is prone to Denial-of-Service attacks. A simple Denial of Service attack on Name Node is enough to halt all the operations of Map Reduce and stop the read-write operation of the Hadoop Distributed file system.

3.4.3. Cross-Site Scripting

Cross-site Scripting [24] is a type of attack where malicious code is injected into any web application that is vulnerable. Cross-Site scripting is different from other attacks such in a way not intended for the implementation in question. But actually, the users of web applications are at risk here. The Cross-site script attacks can be categorized into two types: stored, reflected. Stored attacks also go by the name persistent and are more damaging than the reflected attacks as it is directly injected into the vulnerable web applications. Whereas in reflected, the malicious script is reflected directly onto the user web browser.

3.4.4. Present Attacks

The Hadoop framework due to its open ports and IP address has always been an object of attacks by all the attackers, due to which around 5307 of Hadoop Cluster has been exposed with the vulnerable security settings that attackers use to exploit the framework [25]. There was an online search engine designed to show all the details of the servers and all peripheral devices connected to them over the internet, its name was shodan2. The advantages of shodan2 were that it was possible to recommend any security policy but the disadvantage is that it was used by attackers to exploit the system. To tackle this attack and stop stealing some strategies with high high-security policies must be implemented.

Here, is the following table. 4 that gives a comparative analysis of attacks that had been taken at Hadoop.

Table 4: Comparative analysis of various attacks and challenges faced.

Author	Year	Attacks	Features	Challenges	Description
Bhathal Gurjeet Singh [4]	2019	Impersonation Attacks	Authentication	How to authenticate if the person is actually legit and not impersonated by an attacker.	This type of attack occurs when an attacker tries to impersonate the registered or legitimate authority for accessing the resources.
Bhathal Gurjeet Singh [4]	2019	Denial of Service	Authentication, Authorization	The collection of attacks can be diverse or complex	A Denial of Service is a type of attack where an attacker floods the system with an enormous request which makes the system unable to allocate resources to legitimate users.
Bhathal Gurjeet Singh [4]	2019	Cross-Site Scripting	Authentication, Authorization	Set up some anti triggered methods to avoid hijacking of the user accounts	Cross-site Scripting is a type of attack where malicious code is injected into any web application that is vulnerable.

Fu Xiao [28]	2017	Data Leakage	Confidentiality, authentication, authorization	Avoid leaking, destruction, and corruption of confidential information.	Data Leakage is the unapproved transmission of information from inside an association to an outer objective or beneficiary.
Jose Ancy Sherin [29]	2014	DNS reflection amplification	Confidentiality	Misconfiguration of DNS leads to DDoS	DNS reflection attack is basically a type of Distributed Denial of Service attack.
M Mizukoshi [26]	2019	Distributed Denial of Service	Authentication, Authorization	Manual intervention requirement is too much	A DDoS attack includes different associated web-based gadgets, altogether known as a botnet, which are utilized to overpower an objective site with counterfeit traffic.
Xianqing Yu [27]	2015	Cloud Attacks	Confidentiality	How to avoid misconfiguration, unauthorized access, hijacking	Using the public cloud connection characteristics an attacker can try to hide his breaches.
Bhathal Gurjeet Singh [4]	2019	Port Block Attacks	Access Control	How to avoid overcomplication of the application	Sending of packets to a specific port on the host
Bhathal Gurjeet Singh [4]	2019	SYN Flood Attack	Access Control	How to configure the firewall, setting up an IPS	Repetitive initiation of connection without establishing it to make the server busy.

4. Acknowledgements

This paper and the research behind it were only possible because of the guidance of my guide Dr. Deepak Singh Tomar, Associate Professor at MANIT Bhopal. His attention to detail and helping to keep my work on track from the first encounter.

I would also like to thank my other Supervisor Dr. R K Pateriya, Professor at MANIT Bhopal for their encouragement and guidance in carrying out the project work. I also thank MANIT Bhopal for giving me the opportunity to embark on this project.

5. Conclusion

In this study, an analysis of big data vulnerabilities, security threats, and possible attacks was reviewed for a popular framework like Hadoop. Although it was observed that Hadoop was designed in mind to provide maximum efficiency but with the exponential growth of big data has led to Hadoop being left vulnerable to possible attacks, lack of security policies, mechanisms, proper access control, etc. To make Hadoop a more reliable and secure framework a proper authentication server with authorization and auditing is required. At the same time, some mechanism to ensure data protection will be what an ideal framework would be.

6. References

[1] Lai TL, Yuan H. Stochastic approximation: from statistical origin to big-data, multidisciplinary applications. Statistical Science. 2021 Apr;36(2):291-302.

[2] Li, Yun, Manzhu Yu, Mengchao Xu, Jingchao Yang, Dexuan Sha, Qian Liu, and Chaowei Yang. "Big data and cloud computing." In Manual of Digital Earth, pp. 325-355. Springer, Singapore, 2020.

[3] Oussous, Ahmed, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, and Samir Belfkih. "Big Data technologies: A survey." Journal of King Saud University-Computer and Information Sciences 30, no. 4 (2018): 431-448.

[4] Bhathal, Gurjit Singh, and Amardeep Singh. "Big data: Hadoop framework vulnerabilities, security issues and attacks." Array 1 (2019): 100002.

[5] Brauna, T. D., H. J. Siegelb, N. Beckc, L. L. Bölönid, Albert Muthucumaru Maheswarane, Robertsong IR, Theysh JP, and Yaoi MD. "B., Hensgenj, D. and Freundk, RF, "A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems,"." Journal of Parallel and Distributed Computing 61, no. 6 (2001): 810-837.

[6] Gautam, Akansha, and Indranath Chatterjee. "Big data and cloud computing: A critical review." International Journal of Operations Research and Information Systems (IJORIS) 11, no. 3 (2020): 19-38.

[7] Cai, Xiaojun, Feng Li, Ping Li, Lei Ju, and Zhiping Jia. "SLA-aware energy-efficient scheduling scheme for Hadoop YARN." The Journal of Supercomputing 73, no. 8 (2017): 3526-3546.

[8] Dunn-Rankin, Peter, Gerald A. Knezek, Susan R. Wallace, and Shuqiang Zhang. Scaling methods. Psychology Press, 2014.

[9] Mavridis, Ilias, and Helen Karatza. "Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark." Journal of Systems and Software 125 (2017): 133-151.

[10] Ye, Haina, Xinzhou Cheng, Mingqiang Yuan, Lexi Xu, Jie Gao, and Chen Cheng. "A survey of security and privacy in big data." In 2016 16th international symposium on communications and information technologies (ist), pp. 268-272. IEEE, 2016.

[11] Terzi, Duygu Sinanc, Ramazan Terzi, and Seref Sagiroglu. "A survey on security and privacy issues in big data." In 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), pp. 202-207. IEEE, 2015.

[12] Sharif, Ather, Sarah Cooney, Shengqi Gong, and Drew Vitek. "Current security threats and prevention measures relating to cloud services, Hadoop concurrent processing, and big data." In 2015 IEEE International Conference on Big Data (Big Data), pp. 1865-1870. IEEE, 2015.
[13] Derbeko, Philip, Shlomi Dolev, Ehud Gudes, and Shantanu Sharma. "Security and privacy aspects in MapReduce on clouds: A survey." Computer science review 20 (2016): 1-28.
[14] Bertino, Elisa, and Elena Ferrari. "Big data security and privacy." In A comprehensive guide through the Italian database research over the last 25 years, pp. 425-439. Springer, Cham, 2018.
[15] Zhang, Dongpo. "Big data security and privacy protection." In 8th International Conference on Management and Computer Science (ICMCS 2018), vol. 77, pp. 275-278. Atlantis Press, 2018.
[16] Nelson, Boel, and Tomas Olovsson. "Security and privacy for big data: A systematic literature review." In 2016 IEEE international conference on big data (big data), pp. 3693-3702. IEEE, 2016.
[17] Tao, Hai, Md Zakirul Alam Bhuiyan, Md Arafatur Rahman, Guojun Wang, Tian Wang, Md Manjur Ahmed, and Jing Li. "Economic perspective analysis of protecting big data security and privacy." Future Generation Computer Systems 98 (2019): 660-671.
[18] Erraissi, Allae, and Mouad Banane. "Managing Big Data using Model Driven Engineering: From Big Data Meta-model to Cloudera PSM meta-model." In 2020 International Conference on Decision Aid Sciences and Application (DASA), pp. 1235-1239. IEEE, 2020.
[19] Mitre Corp, "CVE Details", 12 October 2021. [Online]. Available: https://www.cvedetails.com/vendor/45/Apache.html
[20] Salleh, Khairulliza Ahmad, and Lech Janczewski. "Security considerations in big data solutions adoption: Lessons from a case study on a banking institution." Procedia Computer Science 164 (2019): 168-176.
[21] Parmar, Raj R., Sudipta Roy, Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, and Tai-Hoon Kim. "Large-scale encryption in the Hadoop environment: Challenges and solutions." IEEE Access 5 (2017): 7156-7163.
[22] Dahbur, Kamal, Bassil Mohammad, and Ahmad Bisher Tarakji. "A survey of risks, threats and vulnerabilities in cloud computing." In Proceedings of the 2011 International conference on intelligent semantic Web-services and applications, pp. 1-6. 2011.
[23] Gavric, Zeljko, and Dejan Simic. "Overview of DOS attacks on wireless sensor networks and experimental results for simulation of interference attacks." Ingeniería e Investigación 38, no. 1 (2018): 130-138.
[24] Gupta, Shashank, and Brij Bhooshan Gupta. "Cross-Site Scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art." International Journal of System Assurance Engineering and Management 8, no. 1 (2017): 512-530.
[25] Millman, Rene. "Thousands of hadoop clusters still not being secured against attacks." SC Media 10 (2017).
[26] M. Mizukoshi and M. Munetomo, "Distributed denial of services attack protection system with genetic algorithms on Hadoop cluster computing framework," *2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, pp. 1575-1580.
[27] Xianqing Yu, P. Ning and M. A. Vouk, "Enhancing security of Hadoop in a public cloud," *2015 6th International Conference on Information and Communication Systems (ICICS)*, 2015, pp. 38-43
[28] Fu, Xiao, Yun Gao, Bin Luo, Xiaojiang Du, and Mohsen Guizani. "Security threats to Hadoop: data leakage attacks and investigation." IEEE Network 31, no. 2 (2017): 67-71.
[29] Jose, Ancy Sherin, and A. Binu. "Automatic detection and rectification of dns reflection amplification attacks with hadoop mapreduce and chukwa." In 2014 Fourth International Conference on Advances in Computing and Communications, pp. 195-198. IEEE, 2014.