Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

Big data Security and Privacy

Nidhi Gupta[1] , Dr. Savita Shiwani [2]
[1] Research Scholar, Suresh Gyan Vihar University, Jaipur
[2] Professor ,CA, Suresh Gyan Vihar University, Jaipur

Abstract- Big Data is an important issues in the recent years, enables computing resources to be provided as Information Technology services with high efficiency and effectiveness. Big data is a new technological paradigm for data that is generated at high speed, high volume, and with great diversity.On social media thousands of posts are generated per second. Its nature, sharing, storage management, and security and privacy are some crucial issues which are taken up for consideration in this paper. The issues are thoroughly discussed and analysed.

Keywords— Big Data, Hadoop, MapReduce, HDFS, NoSQL, , Security, Privacy

I. Introduction:-

Big data refers to the concept of very large data sets involving three major dimensions or properties named (3Vs). Data can be structured, unstructured, or semi-structured generated either by humans or machines.

First is a volume according to the amount of data located in the storage medium. Second is Variety which refers to the various heterogeneous and complex types of data. The third is velocity which indicates the speed of data processing required to handle that large amount of data.

Most definitions of big data focus on the size of data in storage. Size matters, bu there are other important attributes of big data, namely data variety and data velocity. In addition, each of the three Vs has its own ramifications for analytics.

Eighty percent of data which is generated by social media is unstructured data which cannot be handled by traditional software i.e. DBMS, OLAP etc. Data analysis has been named by many names in 1970 it was named by decision support system, in 1990 it became business intelligence and from 2008 it has now become data analytics. In order to process big data we require other sophisticated tools like Hadoop, R, Hive, NoSQL, search and knowledge discovery, in memory fabric, distributed file system, HDFS, Pig , data virtualization, Polybase, data integration, Sqoop, Presto, etc. The contradictions of big data are identity, transparency, and power [23]. Lambda and kappa are the big data architectures. In order to preserve the privacy, security, threats, vulnerabilities, and attacks some prevention and counter measures are needed. A system is treated secure if is access controlled, integral, authentic, and confidential. Threats and risks are exploited by adversary [24-29].

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

Some policies and mechanisms are framed to prevent, detect, and correct the security attacks on important data. Anti spam, antivirus, firewalls, internet security may be used to thwart the attacks

II.Challenges of Big Data Quality

Big Data can bring cost saving, risk control, improvement of management efficiency, and increment of value into enterprise. In the meanwhile, Big Data brings some challenges:

- Unevenness of Data Quality

The large amount of data. Though the first step of processing data is to gather data, if the gather all data in spite of quality, it is possible to make wrong predictions and decisions. according to view of this condition, after gathering data, it is necessary to select relative data and clean conflicting data.[13]

- Lack of skills

Big data application requires enterprise to design new data analysis models. That's because traditional models are fit to process structured data not big data including multi-type data. Thus, it needs some data science to apply to enterprise data management. The enterprise is short of talents who can design new data analysis models. The talents who not only can design new data analysis models but also know the financial management are fewer. Lack of talents is a severe and long-term issue. Big Data is a sword with two blades. Through affecting the idea, function, mode, and method of financial management, it can bring cost saving, risk control, improvement of management efficiency, and increment of value into enterprise. In the meanwhile, it brings a lot of challenges. Only through fostering strengths and circumvent weaknesses, can an enterprise remain invincible in Big Data era.[13]

III. Big data Features

Figure 1. The six V's of Big Data.

❖ Volume

Refers to the tremendous volume of the data. We usually use TB or above magnitudes to measure this data volume. Velocity means that data are being formed at an unprecedented speed and must be dealt with in a timely manner.

❖ Variety indicates that big data has all kinds of data types, and this diversity divides the data into structured data and unstructured data. These multi typed data need higher data processing capabilities.[14]

❖ The diversity of data sources brings abundant data types and complex data structures and increases the difficulty of data integration

❖ One data type is unstructured

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

data, for example, documents, video, audio, etc. The second type is semi-structured data, including:software packages/modules, spreadsheets, and financial reports. The third is structured data. The quantity of unstructured data occupies more than 80% of the total amount of data in existence.

- ❖ Data change very fast and the "timeliness" of data is very short, which necessitates higher requirements for processing technology

- ❖ Due to the rapid changes in big data, the "timeliness" of some data is very short. If companies can't collect the required data in real time or deal with the data needs over a very long time, then they may obtain outdated and invalid information.

- ❖No unified and approved data quality standards have been formed in China and abroad, and research on the data quality of big data has just begun.

- ❖ In order to guarantee the product quality and improve benefits to enterprises, in 1987 the International Organization for Standardization (ISO) published ISO 9000 standards. Nowadays, there are more than 100 countries and regions all over the world actively carrying out these standards. This implementation promotes mutual understanding among enterprises in domestic and international trade and brings the benefit of eliminating trade barriers. By contrast, the study of data quality standards began in the 1990s.[14]

IV. Big Data Analysis

Big data analytics is differences from traditional analytics Because of the big increase in the volume of data and that led to Many researchers have suggested commercial DBMS and this not suitable with size of data. This type of data is impossible to handle using traditional relational database management systems. New innovative technologies were needed and Google found the solution by using a processing model called MapReduce. There are more solutions to handle Big Data, but the most widely-used one is Hadoop, an open source project based on Google's MapReduce and Google File System. Hadoop was founded by the Apache Software Foundation. The main contributors of the project are Yahoo, Facebook, Citrix, Google, Microsoft, IBM, HP, Cloudera and many others. Hadoop is a distributed batch processing infrastructure which consists of the Hadoop kernel, Hadoop Distributed File System (HDFS), MapReduce and several related projects.[15]

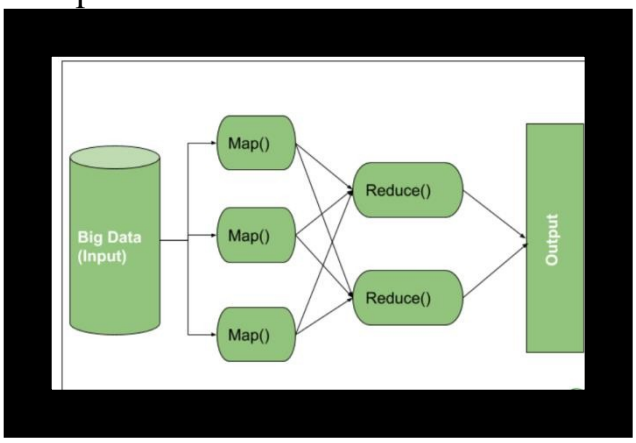
Big Data Analytics refers to the process of collecting, organizing, analyzing large data sets to discover different

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

patterns and other useful information. Big data analytics is a set of technologies and techniques that require new forms of integration to disclose large hidden values from large datasets that are different from the usual ones, more complex, and of a large enormous scale. It mainly focuses on solving new problems or old problems in better and effective ways.

1. MapReduce

MapReduce nothing but just like an Algorithm or a data structure that is based on the YARN framework. The major feature of MapReduce is to perform the distributed processing in parallel in a Hadoop cluster which Makes Hadoop working so fast. When you are dealing with Big Data, serial processing is no more of any use. MapReduce has mainly 2 tasks which are divided phase-wise:
In first phase, Map is utilized and in next phase Reduce is utilized.



Here, we can see that the Input is provided to the Map() function then it's output is used as an input to the Reduce function and after that, we receive our final output. Let's understand What this Map() and Reduce() does.
As we can see that an Input is provided to the Map(), now as we are using Big Data. The Input is a set of Data. The Map() function here breaks this DataBlocks into Tuples that are nothing but a key-value pair. These key-value pairs are now sent as input to the Reduce(). The Reduce() function then combines this broken Tuples or key-value pair based on its Key value and form set of Tuples, and perform some operation like sorting, summation type job, etc. which is then sent to the final Output Node. Finally, the Output is Obtained.
The data processing is always done in Reducer depending upon the business requirement of that industry. This is How First Map() and then Reduce is utilized one by one.

2. HDFS

HDFS(Hadoop Distributed File System) is utilized for storage permission is a Hadoop cluster. It mainly designed for working on commodity Hardware devices(inexpensive devices), working on a distributed file system design. HDFS is designed in such a way that it believes more in storing the data in a large chunk of blocks rather than storing small data blocks.

Hadoop File System (HDFS) is a distributed file system. All types of files can be stored in the Hadoop file system. HDFS in Hadoop provides Fault-tolerance and High availability to

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

the storage layer and the other devices present in that Hadoop cluster. Data storage Nodes in HDFS.

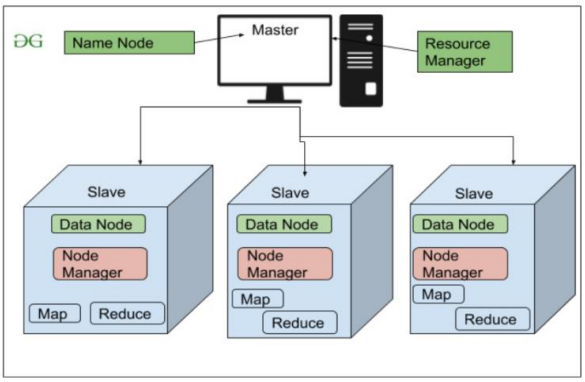
- NameNode(Master)
- DataNode(Slave)

NameNode:NameNode works as a Master in a Hadoop cluster that guides the Datanode(Slaves). Namenode is mainly used for storing the Metadata i.e. the data about the data. Meta Data can be the transaction logs that keep track of the user's activity in a Hadoop cluster.

Meta Data can also be the name of the file, size, and the information about the location(Block number, Block ids) of Datanode that Namenode stores to find the closest DataNode for Faster Communication. Namenode instructs the DataNodes with the operation like delete, create, Replicate, etc.

DataNode: DataNodes works as a Slave DataNodes are mainly utilized for storing the data in a Hadoop cluster, the number of DataNodes can be from 1 to 500 or even more than that. The more number of DataNode, the Hadoop cluster will be able to store more data. So it is advised that the DataNode should have High storing capacity to store a large number of file blocks.

High Level Architecture Of Hadoop



File Block In HDFS: Data in HDFS is always stored in terms of blocks. So the single block of data is divided into multiple blocks of size 128MB which is default and you can also change it manually.

This is because for running Hadoop we are using commodity hardware (inexpensive system hardware) which can be crashed at any time. We are not using the supercomputer for our Hadoop setup. That is why we need such a feature in HDFS which can make copies of that file blocks for backup purposes, this is known as fault tolerance.

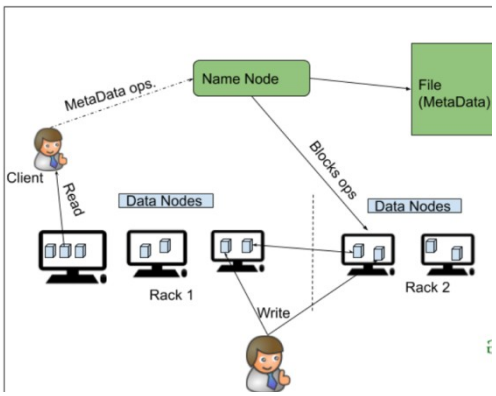
Now one thing we also need to notice that after making so many replica's of our file blocks we are wasting so much of our storage but for the big brand organization the data is very much important than the storage so nobody cares for this extra storage. You can configure the Replication factor in your hdfs-site.xml file.

Rack Awareness The rack is nothing but just the physical collection of nodes in our Hadoop cluster (maybe 30 to 40). A large Hadoop cluster is consists of so many Racks . with the help of this

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

Racks information Namenode chooses the closest Datanode to achieve the maximum performance while performing the read/write information which reduces the Network Traffic.

HDFS Architecture



3. YARN(Yet Another Resource Negotiator)

YARN is a Framework on which MapReduce works. YARN performs 2 operations that are Job scheduling and Resource Management. The Purpose of Job schedular is to divide a big task into small jobs so that each job can be assigned to various slaves in a Hadoop cluster and Processing can be Maximized. Job Scheduler also keeps track of which job is important, which job has more priority, dependencies between the jobs and all the other information like job timing, etc. And the use of Resource Manager is to manage all the resources that are made available for running a Hadoop cluster.

Features of YARN

- Multi-Tenancy
- Scalability
- Cluster-Utilization
- Compatibility

4. Hadoop common or Common Utilities

Hadoop common or Common utilities are nothing but our java library and java files or we can say the java scripts that we need for all the other components present in a Hadoop cluster. these utilities are used by HDFS, YARN, and MapReduce for running the cluster. Hadoop Common verify that Hardware failure in a Hadoop cluster is common so it needs to be solved automatically in software by Hadoop Framework.

V. How we can secure big data:

- **Secure Data Storage and Transaction Logs**

Storage control is a critical component of Big Data reliability. By using signed message digests to have a cryptographic identifier for each digital file or record and to use a technique known as a secure untrusted data repository (SUNDR) to detect unauthorized file modifications by malicious server agents

- **Endpoint Filtering and Validation**

Using a mobile device management solution, you can use trusted

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

credentials, perform resource verification, and link only trusted devices to the network. Using statistical similarity detection and outlier detection strategies, you can process malicious inputs while defending against Sybil attacks (one person posing as several identities) and ID-spoofing attacks.

- **Real-Time Compliance and Security Monitoring**

Organizations can use techniques like Kerberos, safe shell, and internet protocol protection to get a grip on real-time data by using Big Data analytics. It's then simple to monitoring logs, set up front-end security mechanisms like routers and server-level firewalls, and start putting security controls in place at the cloud, network, and application levels.

Graph Databases uses graph architecture for semantic inquiry with nodes, edges, and properties to represent and store data. Role of Graph Databases in Big Data Analytics

- **Preserve Data Privacy**

Employee awareness training centers on new privacy laws and ensures that information technology is kept up to date by using authorization processes. In addition, data leakage from different databases can be regulated by analyzing and tracking the infrastructure that connects the databases.

- **Big Data Cryptography**

Mathematical cryptography has improved significantly. Enterprises can run Boolean queries on encrypted data by creating a method to scan and filter encrypted data, such as the searchable symmetric encryption (SSE) protocol.

- **Granular Access Control**

The two main aspects of access management are limiting and allowing user access. The key is to create and execute a policy that automatically selects the best option in any given situation.

To set up granular access controls:

1. Immutable elements should be denormalized, and mutable elements should be normalized.
2. Please keep track of confidentiality provisions to make sure they're followed.
3. Keep track of control marks.
4. Keep track of administrative information.
5. To ensure proper data federation, use a single sign-on (SSO) and a labeling system.

Strategies for Granular Access Control, some are listed below :

1. Point out mutable elements and immutable elements.
2. Access labels should be maintained, track admin data too.
3. Use single sign-on, and maintain a proper labeling scheme.
4. Perform audit layer/orchestrator

- **Granular Auditing**

In Big Data protection, granular auditing is essential, particularly after a system attack. Organizations should develop a unified audit view following an attack and include a complete audit

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

trail with quick access to the data to reduce incident response time.
The integrity and security of audit records are also important. Audit data should be kept isolated from other data and safeguarded with granular user access controls and routine reporting. When configuring auditing, keep Big Data and audit data separate, and allow all necessary logging. An orchestrator tool like ElasticSearch can make it easier to do.

- **Data Provenance**

It's provenance metadata that Big Data applications produce. This is a different kind of data that requires special protection. Creating an infrastructure authentication protocol that manages access and sets up daily status alerts, and constantly checks data integrity with checksums.

6. Conclusions

Big data are not safe. The aim of Big Data analytics for safety is to obtain information that can be activated in real time. While Big Data analytics have a lot of promise, they still have a way to achieve full potential. Numerous security procedures were submitted for Big Data Analytics. A particular protocol should be used because of the safety issues and the application. Large data analytics focus on security and privacy issues and enhance the safety and pry- vacy of Big Data platforms.We've identified a number of security and privacy issues that Big Data technologies should consider in this article. We also discussed some potential solutions and strategies for protecting this distributed system. We also address few privacy violations and also discussed encryption algorithms used in Big Data analytics. Some of these protective mea- sures will be included in an open source Big Data analysis tool as part of future development. Currently, several privacy-preserving strategies for Big Data exist, such as anony- mization protection technology, access control, encryption, unstructured distribution, data tracing, differential privacy protection, anonymization, and so on. We end with a few re commendations for improving the efficiency of a Big Data project, and provide secure possible techniques and proposed solutions and model that minimizes privacy violation showing four different types of data protection violations and the involvement of different entities in reducing their impacts. However, in algorithms as well as in system areas, further research is needed to deal with the increasingly many problems ahead.

VII. References

Abdulhamid, S. M., Abd Latiff, M. S., Chiroma, H., Osho, O., Abdul- Salaam, G., Abubakar, A. I., & Herawan, T. (2017). A review on mobile SMS spam filtering techniques. *IEEE Access*, *5*, 15650–15666.
https://doi.org/10.1109/ACCESS.2017.2666785

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: Preserving security and privacy. *Journal of Big Data*, *5*(1), 1–18. https://doi.org/10.1186/s40537-017-0110-7

Adjei, J. K., Adams, S., Mensah, I. K., Tobbin, P. E., & Odei- Appiah, S. (2020). Digital identity management on social media: Exploring the factors that influence personal informa- tion disclosure on social media. *Sustainability (Switzerland)*, *12*(23), 1–17. https://doi.org/10.3390/su12239994

Aftab, M. O., Javed Awan, M., Khalid, S., Javed, R., & Shabir,

H. (2021). *Executing spark BigDL for leukemia detec- tion from microscopic images using transfer learning* [Conference session]. 2021 1st International Conference on Artificial Intelligence and Data Analytics, CAIDA 2021, Riyadh, Saudi Arabia, pp. 216–220. https://doi.org/10.1109/ CAIDA51941.2021.9425264

Ahmed, H. M., Awan, M. J., Khan, N. S., Yasin, A., & Shehzad, H.

M. F. (2021). Sentiment analysis of online food reviews using Big Data analytics. *Ilkogretim Online*, *20*(2), 827–836. https:// doi.org/10.17051/ilkonline.2021.02.93

Alam, T. M., & Awan, M. J. (2018). Domain analysis of infor- mation extraction techniques. *International Journal of Multidisciplinary Sciences and Engineering*, *9*(6), 1–9.

Alguliyev, R., & Imamverdiyev, Y. (2014). *Big Data: Big prom- ises for information security* [Conference session]. 8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 – Conference Proceedings, Astana, Kazakhstan. https://doi.org/10.1109/ ICAICT.2014.7035946

Ali, Y., Farooq, A., Alam, T. M., Farooq, M. S., Awan, M. J., & Baig, T. I. (2019). Detection of schistosomiasis factors using association rule mining. *IEEE Access*, *7*, 186108–186114. https://doi.org/10.1109/ACCESS.2019.2956020

Alier, M., Jose, M., Guerrero, C., Amo, D., Severance, C., & Fonseca, D. (2021). *Privacy and E-learning : A pending task. Sustainability*, *13*(16), 9206.

Alshboul, Y., Wang, Y., & Nepali, R. K. (2015). Big Data life- cycle: Threats and security model [Conference session]. 2015 Americas Conference on Information Systems, AMCIS 2015, Fajardo, Puerto Rico, pp. 1–7.

Ambalavanan, V. (2020). Cyber threats detection and mitigation using machine learning. In P. Ganapathi & D. Shanmugapriya (Eds.), *Handbook of research on machine and deep learning applications for cyber security* (pp. 132–149). IGI Global.

Anam, M., Ponnusamy, V., Hussain, M., Nadeem, M. W., Javed, M., Goh, H. G., & Qadeer, S. (2021). Osteoporosis predic- tion for trabecular bone using machine learning: A review. *Computers, Materials and Continua*, *67*(1), 89–105. https:// doi.org/10.32604/cmc.2021.013159

Application, F., Data, P., Examiner, P., & Andrews, M. (1999).

United States Patent (19). United States Patent.

Applications, C., Technology, I., Engineering, S., Engineering, S., & Engineering, C. (n.d.). *Efficient Residential Load Forecasting using Deep Learning Approach Rida Mubashar*

* *Mazhar Javed Awan Muhammad Ahsan Awais Yasin Vishwa Pratab Singh. X*(2006). United States Patent.

Aradau, C., & Blanke, T. (2015). The (Big) data-security assem- blage: Knowledge and critique. *Big Data and Society*, *2*(2), 1–12. https://doi.org/10.1177/2053951715609066

Awan, M. J. (2020). Fake news classification bimodal using convo- lutional neural network and long short-term memory. *Article in International Journal of Emerging Technologies in Learning (IJET)*, *11*(5), 209–212.

Awan, M. J., Khan, M. A., Ansari, Z. K., Yasin, A., & Shehzad, H.

M. F. (forthcoming). Fake profile recognition using big data analytics in social media platforms. *International Journal of Computer Applications in Technology*.

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

https://www.inder-science.com/info/ingeneral/forthcoming.php?jcode=ijcat

Awan, M. J., Khan, R. A., Nobanee, H., Yasin, A., Anwar, S. M., Naseem, U., & Singh, V. P. (2021). A recommenda- tion engine for predicting movie ratings using a Big Data approach. *Electronics (Switzerland)*, *10*(10),1215.https://doi.

Awan, M. J., Rahim, M. S. M., Nobanee, H., Munawar, A., Yasin, A., & Zain, A. M. (2021). Social media and stock market prediction: A Big Data approach. *Computers, Materials and Continua*, *67*(2), 2569–2583. https://doi.org/10.32604/ cmc.2021.014253

Awan, M. J., Rahim, M. S. M., Nobanee, H., Yasin, A., Khalaf, O. I., & Ishfaq, U. (2021). A Big Data approach to black Friday sales. *Intelligent Automation and Soft Computing*, *27*(3), 785– 797. https://doi.org/10.32604/iasc.2021.014216

Awan, M. J., Rahim, M. S. M., Salim, N., Ismail, A. W., & Shabbin,

H.(2019). Acceleration of knee MRI cancellous bone classifi- cation on google colaboratory using convolutional neural net-work. *International Journal of Advanced Trends in Computer Science and Engineering,* *8*(1.6 Special Issue), 83–88. https:// doi.org/10.30534/ijatcse/2019/1381.62019

Awan, M. J., Rahim, M. S. M., Salim, N., Mohammed, M. A., Garcia-Zapirain, B., & Abdulkareem, K. H. (2021). Efficient detection of knee anterior cruciate ligament from magnetic resonance imaging using deep learning approach. *Diagnostics*, *11*(1), 105. https://doi.org/10.3390/diagnostics

Awan, M. J., Raza, A., Yasin, A., Muhammad, H., & Shehzad, F. (2021). The customized convolutional neural network of face

emotion expression classification. *Annals of R.S.C.B*, *25*(6), 5296–5304.

Barth-Jones, D. C. (2012). The "Re-Identification" of governor William Weld's medical information: A critical re-examina- tion of health data identification risks and privacy protections, then and now. https://doi.org/10.2139/ssrn.2076397

Battams, K. (2015). *Stream mining for solar physics: Applications and implications for big solar data* [Conference Session]. Proceedings – 2014 IEEE International Conference on Big Data, IEEE Big Data 2014, Washington, DC, pp.18–26. https://doi.org/10.1109/BigData.2014.7004400

Butpheng, C., Yeh, K. H., & Xiong, H. (2020). Security and privacy in IoT-cloud-based e-health systems-A comprehensive review. *Symmetry*, *12*(7), 1–35. https://doi.org/10.3390/sym12071191

Cárdenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy*, *11*(6), 74–76. Chandramouli, B., Goldstein, J., & Duan, S. (2012). *Temporal ana- lytics on Big Data for web advertising* [Conference Session].

Proceedings – International Conference on Data Engineering, Arlington, VA, pp. 90–101. https://doi.org/10.1109/ ICDE.2012.55

Chandrasekar, Dr. C. (2018). Classification techniques using spam filtering email. *International Journal of Advanced Research in Computer Science*, *9*(2), 402–410. https://doi.org/10.26483/ ijarcs.v9i2.5571

Che, D., Safran, M., & Peng, Z. (2013). From Big Data to Big Data mining: Challenges, issues, and opportunities. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7827 LNCS*, pp. 1–15. https://doi.org/10.1007/978-3-642-40270-8_1

Chen, X. W., & Lin, X. (2014). Big Data deep learning: Challenges and perspectives. *IEEE Access*, *2*,514–525.https://doi. org/10.1109/ACCESS.2014.2325029

Cheng, L., Liu, F., & Yao, D. D. (2017). Enterprise data breach: Causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and*

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

Knowledge Discovery, *7*(5), 1–14. https://doi.org/10.1002/widm.1211

Colesky, M., Hoepman, J. H., & Hillen, C. (2016). A critical anal- ysis of privacy design strategies. Proceedings – 2016 IEEE Symposium on Security and Privacy Workshops, SPW 2016, San Jose, CA, pp. 33–40. https://doi.org/10.1109/SPW.2016.23

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, *4*(10), 13–21. https://doi.org/10.22215/timreview835

Csányi, G. M., Nagy, D., Vági, R., Vadász, J. P., & Orosz, T. (2021). Challenges and open problems of legal document ano- nymization. *Symmetry*, *13*(8), 1–25. https://doi.org/10.3390/ sym13081490

De Goede, M. (2014). The politics of privacy in the age of preemp- tive security. *International Political Sociology*, *8*(1), 100–104. https://doi.org/10.1111/ips.12042

Dev Mishra, A., & Beer Singh, Y. (2017). *Big Data analytics for security and privacy challenges* [Conference session]. Proceeding – IEEE International Conference on Computing, Communication and Automation, ICCCA 2016, Greater Noida, India, pp. 50–53. https://doi.org/10.1109/CCAA.2016.781368 8 Dumitras, T., & Shou, D. (2011). *Toward a standard benchmark for computer security research: The worldwide intelligence net- work environment (WINE)* [Conference session]. Proceedings of

the 1st Workshop on Building Analysis Datasets and GatheringExperience Returns for Security, Salzburg, Austria, BADGERS 2011, pp. 89–96. https://doi.org/10.1145/1978672.1978683

Ebert, I., Wildhaber, I., & Adams-Prassl, J. (2021). Big Data in the workplace: Privacy due diligence as a human rights- based approach to employee privacy protection. *Big Data and Society*, *8*(1). https://doi.org/10.1177/20539517211013051

Farkas, C. (2014). *Big Data analytics: Privacy protection using semantic web technologies* [Conference session]. NSF Workshop on Big Data Security and Privacy, Texas, San Antonio, United States.

Firdausi, I., Lim, C., Erwin, A., & Nugroho, A. S. (2010). *Analysis of machine learning techniques used in behavior-based mal- ware detection* [Conference session]. Proceedings – 2010 2nd International Conference on Advances in Computing, Control and Telecommunication Technologies, ACT 2010, Jakarta, Indonesia, pp. 201–203. https://doi.org/10.1109/ACT.2010.33

Florea, D., & Florea, S. (2020). Big Data and the ethical implica- tions of data privacy in higher education research. *Sustainability (Switzerland)*, *12*(20), 1–11. https://doi.org/10.3390/su1220

François, J., Wang, S., Bronzi, W., State, R., & Engel, T. (2011). *BotCloud: Detecting botnets using MapReduce* [Conference session]. 2011 IEEE International Workshop on Information Forensics and Security, WIFS 2011, Iguacu Falls, Brazil. https://doi.org/10.1109/WIFS.2011.6123125

Gahi, Y., & Alaoui, I. El. (2019). A secure multi-user database- as-a-service approach for cloud computing privacy. *Procedia Computer Science*, *160*, 811–818. https://doi.org/10.1016/j. procs.2019.11.006

Gai, K., Qiu, M., & Zhao, H. (2016). *Security-aware efficient mass distributed storage approach for cloud systems in Big Data* [Conference session]. Proceedings – 2nd IEEE International Conference on Big Data Security on Cloud, IEEE BigDataSecurity 2016, 2nd IEEE International Conference on High Performance and Smart Computing, IEEE HPSC 2016 and IEEE International Conference on Intelligent Data and S, New York, NY, pp. 140–145. https://doi.org/10.1109/ BigDataSecurity-HPSC-IDS.2016.68

Geist, A., & Reed, D. A. (2017). A survey of high-performance computing scaling challenges. *The International Journal of High Performance Computing Applications*, *31*(1),

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

104–113. https://doi.org/10.1177/1094342015597083

Guo, J., Yang, M., & Wan, B. (2021). A practical privacy-preserv- ing publishing mechanism based on personalized k-anonymity and temporal differential privacy for wearable iot applications. *Symmetry*, *13*(6), 1043. https://doi.org/10.3390/sym13061043

Gupta, M., Jain, R., Arora, S., Gupta, A., Awan, M. J., Chaudhary, G., & Nobanee, H. (2021). AI-enabled COVID-19 outbreak analysis and prediction: Indian states vs. union territories. *Computers, Materials and Continua*, *67*(1), 933–950. https:// doi.org/10.32604/cmc.2021.014221

Gurajala, S., White, J. S., Hudson, B., & Matthews, J. N. (2015, July). Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach [Conference session]. Proceedings of the 2015 International Conference on Social Media & Society, torpnto, ON, Canada.

Inbarani, H. H., & Kumar, S. S. (2015). *Big Data in complex systems*(*Vol. 9*). Springer. https://doi.org/10.1007/978-3-319-11056-1

International Standard Organization. (2011). *International stan-*

dard ISO/IEC information technology— Security techniques— Application security.Jacobs, B., & Popma, J. (2019). Medical research, Big Data and the need for privacy by design. *Big Data and Society*, *6*(1), 1–5. https://doi.org/10.1177/205395171882435 2

Javed, R., Saba, T., Humdullah, S., Mohd Jamail, N. S., & Javed Awan, M. (2021). *An efficient pattern recognition based method for drug-drug interaction diagnosis* [Conference session]. 2021 1st International Conference on Artificial Intelligence and Data Analytics, CAIDA 2021, Riyadh, Saudi Arabia, pp. 221–226. https://doi.org/10.1109/CAIDA51941.2021.942 5062

Joseph, A. D., Nelson, B., Nelson, B., & Tygar, J. D. (2019). *Adversarial Machine Learning*. Cambridge University Press. https://doi.org/10.1017/9781107338548

Jusas, V., Japertas, S., Baksys, T., & Bhandari, S. (2019). Logical filter approach for early stage cyber-attack detection. *Computer Science and Information Systems*, *16*(2), 491–514. https://doi. org/10.2298/CSIS190122008J

Jusas, V., & Samuvel, S. G. (2019). Classification of motor imagery using combination of feature extraction and reduction methods for brain-computer interface. *Information Technology and Control*, *48*(2), 225–234. https://doi.org/10.5755/j01.itc.48.2.23091

Kantarcioglu, M., & Shaon, F. (2019). *Securing Big Data in the age of AI* [Conference session]. Proceedings – 1st IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2019, Los Angeles, CA, pp. 218–220. https://doi.org/10.1109/TPS-ISA48467.2019.00035

Khan, N., Naim, A., Hussain, M. R., Naveed, Q. N., Ahmad, N., & Qamar, S. (2019). *The 51 V's of Big Data: Survey, tech- nologies, characteristics, opportunities, issues and chal-lenges* [Conference session]. ACM International Conference Proceeding Series, Crete, Greece, Part F1481, pp. 19–24. https://doi.org/10.1145/3312614.3312623

Kim, J., & Park, N. (2020). A face image virtualization mechanism for privacy intrusion prevention in healthcare video surveil- lance systems. *Symmetry*, *12*(6), 891. https://doi.org/10.3390/ SYM12060891

Koo, J., Kang, G., & Kim, Y. G. (2020). Security and privacy in Big Data life cycle: A survey and open challenges. *Sustainability (Switzerland)*, *12*(24), 1–32. https://doi.org/10.3390/su122410571 Krishna, R. R., Priyadarshini, A., Jha, A. V, Appasani, B., & Srinivasulu, A. (2021). State-of-the-art review on IoT threats and attacks : Taxonomy, challenges and solutions. *Sustainability*,

Kuhn, D. R., Walsh, T. J., & Fries, S. (2005). Security consid- erations for voice over IP

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

systems recommendations of the national institute of standards and technology. *National Institute of Standards and Technology, 800–58*, 1–93.

Lee, W., Stolfo, S. J., Chan, P. K., Eskin, E., Fan, W., Miller, M., Hershkop, S., & Zhang, J. (2001). *Real time data mining- based intrusion detection* [Conference session]. Proceedings – DARPA Information Survivability Conference and Exposition II, Anaheim, CA, DISCEX 2001, *1*, pp. 89–100. https://doi. org/10.1109/DISCEX.2001.932195

Li, M., Zang, W., Bai, K., Yu, M., & Liu, P. (2013). *MyCloud – Supporting user-configured privacy protection in cloud com- puting* [Conference session]. ACM International Conference Proceeding Series, New Orleans, LA, United States, pp. 59–68. https://doi.org/10.1145/2523649.2523680

Liu, Z. C., Xiong, L., Peng, T., Peng, D. Y., & Liang, H. B. (2018). A realistic distributed conditional privacy-preserveing authentication scheme for vehicular ad hoc networks. *IEEE Access*, *6*, 26307–26317. https://doi.org/10.1109/ ACCESS.2018.2834224

Manjula, K., & Anandaraju, M. B. (2018). A comparative study on feature extraction and classification of mind waves for brain computer interface (BCI). *International Journal of Engineering and Technology(UAE)*, *7*(1), 132–136. https://doi. org/10.14419/ijet.v7i1.9.9749

McDermott, Y. (2017). Conceptualising the right to data protection in an era of Big Data. *Big Data and Society*, *4*(1), 1–7. https:// doi.org/10.1177/2053951716686994

Mohan, K., Shrivastva, P., Rizvi, M. A., & Singh, S. (2014). *Big Data privacy based on differential privacy a hope for Big Data*. https://doi.org/10.1109/.167

Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. *Applied Sciences (Switzerland)*, *11*(9), 4164. https://doi.org/10.3390/app11094164

Nair, K. K., Helberg, A., & Van Der Merwe, J. (2016). *An approach to improve the match-on-card fingerprint authentication system security* [Conference session]. 2016 6th International Conference on Digital Information and Communication Technology and Its Applications, Konya, Turkey, DICTAP 2016, pp. 119–125. https:// doi.org/10.1109/DICTAP.2016.7544012

Ninghui, L., Tiancheng, L., & Venkatasubramanian, S. (2007). *T-Closeness: Privacy beyond k-anonymity and ℓ-diversity* [Conference session]. Proceedings – International Conference on Data Engineering, 3, pp. 106–115, Istanbul, Turkey. https:// doi.org/10.1109/ICDE.2007.367856

Onan, A. (2015). A fuzzy-rough nearest neighbor classifier com- bined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer. *Expert Systems with Applications*, *42*(20), 6844–6852. https://doi. org/10.1016/j.eswa.2015.05.006

Onan, A. (2019). Topic-enriched word embeddings for sarcasm identification. *Advances in Intelligent Systems and Computing*, *984*, 293–304. https://doi.org/10.1007/978-3-030-19807-7_29

Onan, A. (2021). Ensemble of classifiers and term weighting schemes for sentiment analysis in Turkish. *Scientific Research Communications*, *1*(1), 1–12. https://doi.org/10.52460/src.2021.004

Onan, A., & Korukoğlu, S. (2016). Exploring performance of instance selection methods in text sentiment classification. *Advances in Intelligent Systems and Computing*, *464*, 167–179. https://doi.org/10.1007/978-3-319-33625-1_16

Onan, A., & Korukoğlu, S. (2017). A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science*, *43*(1), 25–38. https://doi. org/10.1177/0165551515613226

Onan, A., Korukoğlu, S., & Bulut, H. (2017). A hybrid ensemble pruning approach based on

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 9, Issue 1, 2023, pp 24-37
ISSN: 2395-0196

consensus clustering and multi- objective evolutionary algorithm for sentiment classification. *Information Processing & Management*, *53*(4), 814–833. https://doi.org/10.1016/J.IPM.2017.02.008

Onan, A., & Tocoglu, M. A. (2020). Satire identification in Turkish news articles based on ensemble of classifiers. *Turkish Journal of Electrical Engineering and Computer Sciences*, *28*(2), 1086–1106. https://doi.org/10.3906/elk-1907-11

Onan, A., & Tocoglu, M. A. (2021). A term weighted neural lan- guage model and stacked bidirectional LSTM based framefor sarcasm identification. *IEEE Access*, *9*, 7701–7722. https:// doi.org/10.1109/ACCESS.2021.3049734

Patel, S. C., Graham, J. H., & Ralston, P. A. S. (2008). Quantitatively assessing the vulnerability of critical information sys- tems: A new method for evaluating security enhancements. *International Journal of Information Management*, *28*(6), 483–491. https://doi.org/10.1016/j.ijinfomgt.2008.01.009

Peter, S. (2005). *Ripped by AaLl86. In security*.

Pham, V., & Dang, T. (2019, December). *CVExplorer: Multidimensional visualization for common vulnerabilities and exposures* [Conference session]. Proceedings – 2018 IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, pp. 1296–1301. https://doi.org/10.1109/ BigData.2018.8622092

Pötzsch, S. (2009). Privacy awareness: A means to solve the privacy paradox? *IFIP Advances in Information and Communication Technology*, *298*(216483), 226–236. https:// doi.org/10.1007/978-3-642-03315-5_17

Rajan, S., van Ginkel, W., & Sundaresan, N. (2012, November). Cloud security alliance (CSA): Top ten Big Data security and privacy challenges. *Csa*, *1*, 1–11.

Rastogi, N., Singh, S. K., & Singh, P. K. (2018, November 1). *Privacy and security issues in Big Data: Through Indian prospective* [Conference session]. Proceedings - 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages, IoT-SIU 2018, Bhimtal, India. https:// doi.org/10.1109/IoT-SIU.2018.8519858

Sánchez-Moreno, D., Batista, V. L., Vicente, M. D. M., Lázaro, Á. L. S., & Moreno-García, M. N. (2020). Exploiting the user social context to address neighborhood