

Hadoop Framework for Big Data Analytics to Analyze Attacks on Business Data



Siva Sankara Reddy Donthi Reddy, Udaya Kumar Ramanadham

Abstract: *Big Data consist large volumes of data sets with various formats i.e., structured, unstructured and semi structured. Big Data requires security because day by day attackers attack on it in different manner. Big Data Security Analytics analyses Big Data for finding various threats and complex attacks. By increasing the number of targeting attacks on data and one side rapid growing of data, it is too difficult to analyze accurately. The Security Analytics Systems are used the untrusted data. So, strong security analytical tools are required to analyze the data. The organizations and industries exchange the data through networks dynamically, so this may become more vulnerable for data misusing and theft. Attackers are more advanced in the attacking on data that the existing security mechanisms are not identified before damaging. At present, the collecting and analyzing various attacks is major challenging task for Security Analytics Systems, to take suitable decision. In this research paper, we have addressed about Hadoop tool that how it analyses Big Data and how Big Data Security Analytics is applied to analyze the various threats and securing the business data more accurately.*

Keywords: *Big Data, Hadoop, MapReduce, Big Data Security Analytics, Threats.*

I. INTRODUCTION

With the enhancement of Internet and technology more rapidly, all computers are connected to communicate themselves either using fixed networking or mobile communication. Daily huge data is producing by various users in unstructured form through communication media. So, these data is not manageable and more challengeable task [1]. The main focus is showed on gathering the unstructured form data from different sources and it will be processed the data to get structured form from unstructured form, so this data can be accessed easily [2]. By using Big Data Analytic tools, data can be analyzed and processed to transform into more understandable and meaningful form. Big Data Analytic tools are received huge data and unstructured form as input such as videos, images, audio, web-pages, texts or e-mails etc. A simple algorithm is implemented in Big Data Analytics for large data sets without compromising its performance. Big Data Analytics can use the various tools for analyzing and processing the data in the meaningful manner.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Siva Sankara Reddy Donthi Reddy*, PhD scholar, Department of CSE, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamilnadu, India, d.sivasankarareddy@gmail.com

Udaya Kumar Ramanadham, Professor, Department of Information Technology, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamilnadu, India, rsukumar2007@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Hadoop is used to improve the performance of data processing [3]. Day by day, Hackers are becoming stronger with various attacks on processed information which is available on internet. Hackers create threats on analyzed data [20]. The Big Data Security Analytical tools are used to detect vulnerabilities and intrusions, and to gather and analyze security data [5]. The primary aim of using Big Data techniques is to analyze data threatening and applying data security mechanisms on data. The organizations chose the different types of hosts with a range of Security Analytics Sources (SAS) to get data. The Big Data Security Analytics are used to generate alerts and relevant messages. These messages are transmitted to trusted machine to analyze and to initiate action accordingly. It is like as Host based Intrusion Detection System (HIDS) and also as antivirus system which can write a syslog. The actions are initiated based on reports of HIDS. e.g., Security and Information Event Monitoring (SIEM) system. The targeted attacks and malicious code are key aspects for organizations and government. The Advanced Persistent Threats (APT) is a subset of threats which is well trained adversaries that attacks on targeted highly sensitive data [25].

II. LITERATURE REVIEW

Mohan Shivaji Kamble et al. [1] focused on the massive face identification system for recognizing a person face among all the persons faces in the public places. In this system, two techniques uses to identify the face: the first one is 3D face detecting technique and the second one is Hadoop detection technique. These techniques detect the similar kind of face by 3D face features from huge large sets of data of person faces. Hadoop process the operations parallel and it increase computation speed. The algorithm produces accurate results and it is an efficient for large sets of data.

K.Tamilselvi et al. [2] presented some problems in Big Data Paradigm and the implementation of Big Data techniques in organization to meet their requirements. In this paper, the authors mainly described the Hadoop, HDFS, MapReduce. This tools can produce the good results with high accuracy by overcome the demerits.

Shaili et al. [3] presented the data analysis which is the combination of concepts, tools and algorithms of machine learning and statistics to analyze huge data sets for gaining insight, understanding and effective knowledge. The authors described the MapReduce applications for optimizing the energy and the Apache Framework was used in distributed processing environment across all the clusters of systems for large data sets using familiar programming methods.

Hadoop Framework for Big Data Analytics to Analyze Attacks on Business Data

Suruchi Padhy et al. [4] suggested different techniques to solve the issues and problems using Map Reduce framework under Hadoop Distributed File System (HDFS). The authors described minimized mapreduce technique for file indexing usage with mapping, sorting, shuffling and finally reducing. In this paper, the Map Reduce techniques were studied to implement for Big Data analysis using HDFS.

K.Venkatesh et al. [5] studied data mining techniques for big data, distributed frameworks and the security and privacy preserving techniques for big data which is being accumulated in enterprises. The authors presented frameworks such as Hadoop, Halopop, Sailfish, Dryad and AROM. It is essential to ensure the privacy of the individuals that are associated with data being mined during data processing in distributed environment like cloud computing. The security and privacy preserving techniques such as data mining for big data, distributed programming frameworks and the PPDM techniques were presented.

Nishant Rajput et al. [6] presented Hadoop MapReduce frameworks and HDFS. These can be used to process huge and complex data sets. Hadoop collaborates the MapReduce model which runs in the private cloud. Hadoop can have ability to empower the clients for making the clustered item servers. MapReduce is a free platform layer which is necessity for cloud suppliers. It can empower the clients to perform the big data processing and investigating operations.

Vaishnavi Chillal et al. [7] proposed framework that it is designed for distributed storage which is exceptionally useful for checking enormous information. This framework is supported for E-banking exchange operations. The authors were implemented Hadoop framework for performing analytical operations on E-Banking data.

Prof. Girija Chiddarwar et al. [8] developed method which is a Hadoop based framework for processing large sets of data. The authors used various data mining techniques for extracting useful information such as clustering with connection, clustering with center. After analysis of the data by using the data mining techniques, the security threats can detect using the results in the network. It can also be used for informing to the administrator which may aware of detection of intrusion, sql injections, system errors, etc.

Goddilla Nagarjuna Reddy et al. [9] designed a model which includes Hadoop and MapReduce, HDFS for storing and processing the big data of Online Retail Domain. In this research work, the designed model demonstrates fundamentally builds the execution of multifaceted nature examination.

L. Greeshma et al. [10] designed architecture for Hadoop Distributed File System and MapReduce to store and accessing the information from data sets. In this research work, the authors implemented a MapReduce based WordCount application under object oriented programming paradigm. The input file is divided as small chunks or units or splits or tokens by using StringTokenizer class in Java. The output file represents as <<key>, value>. The larger input files are uploaded for testing the process to get accurate results and the performance of Hadoop framework is evaluated. After conducting tests on this model, the authors have achieved good performance with larger number of files. The read-write operations are achieved by map and reduce tasks.

Saloni Minocha et al. [11] presented HadoopWEB framework It is a middleware platform for web services between the end-user and Virtual Hadoop Cluster. It can be used to process MapReduce jobs by users. By using HadoopWeb, some experiments were conducted for various MapReduce applications on Chicago's Crimes real time data. The HadoopWEB is used for executing the MapReduce jobs.

Dr. Urmila R. Pol [12] presented the concepts of Big Data Analytics using Hadoop and MapReduce. The Hadoop is free and open-source software for processing the tasks in reliable and scalable in the distributed computing environment. The authors were discussed the installation process of Hadoop, Hadoop MapReduce framework.

Dr. Urmila R. Pol [13] presented the concepts of Big Data Technologies and analysis process using Hadoop mapreduce, pig and hive. By using simple models of programming, the Hadoop framework allows the data sets in distributed processing manner across the clusters of computers. Hadoop MapReduce provides low level of abstraction but Hive and Pig technologies are provided high level abstraction. The authors discussed the comparison of Big Data Analytical like Pig, Hive, Hadoop MapReduce on performance.

M. Dhavapriya et al. [14] discussed problems, issues and solutions of Big Table, Hadoop, Map Reduce. These tools are used for storing and managing BigData. The authors were suggested the various methods for processing the operations efficiently using MapReduce framework on Hadoop Distributed File System (HDFS). It minimizes the additional resources for making the file indexing such as mapping, sorting, shuffling and reducing.

Amol Bansod [15] proposed Spark framework for reliable and effective analysis of big data in HDFS. It increases the speed of computation of iterative algorithms and completes them with in limited period of time as compared to Hadoop MapReduce. Apache Spark also provides a high performance, highly scalable and fault tolerant framework for big data analysis

Joelson A. dos Santos et al. [16] proposed parallelize Single-Linkage clustering framework using MapReduce. This method shows the efficiency and good scaling when it is evaluated in terms of both runtime and quality on a number of datasets.

Rotsnarani Sethy et al. [17] discussed Hadoop tool framework for solving the Big Data problems and issues. It can be used for data analyzing mainly. The authors presented hadoop related tools that these works in the distributed environments for processing the large data sets. The authors were presented the Big Data concepts and to implement applications in distributed environment using Hadoop using MapReduce algorithm and to handle node failures.

Rahul Beakta [18] discussed the Big Data concepts, advantages of Big Data, future scope for Big Data application implementation, present opportunities of Big Data and challenges. The authors also presented an overview of opportunities in healthcare domain and describes Hadoop tool concepts which are used in the applications of Big Data in Data Mining.

Harshawardhan S. Bhosale et al. [19] focused mainly different issues, problems and the technical challenges of Big Data. The technical challenges may include scale issues, quality, no proper structure, No proper mechanism for error-free, privacy issues. The authors described Hadoop tool that it is free for downloading and using to process on Big Data.

Bhawna Gupta et al. [20] discussed Big Data technique Hadoop which is used to analyze big data. The authors also discussed security analytics for big data and so these techniques are very important to deny the threats to provide reliable security for an organization data more effectively.

Dr. Siddaraju et al. [21] suggested different methods to solve the problems of big data using MapReduce framework under the HDFS. The authors also studied MapReduce technique. It is necessity to implement analysis by using HDFS. The authors were exploited the MapReduce structure for resolving the critical issues with good analysis of Big Data, challenges and issues of higher scale datasets in various domains. MapReduce provides easy and basic mechanism for scaling applications.

Vidyullatha Pellakuri et al. [22] discussed Hadoop MapReduce framework which is a broad scale, open source software framework devoted to scalable, distributed, data-intensive computing. The large data sets are broken into small equalized chunks by using MapReduce framework and also process the scheduling. MapReduce can support for more no of nodes and huge sets of data and it is fault tolerant.

Rabi Prasad Padhy [23] discussed the overall architecture of Hadoop, The various implementations and application developments are carried out Cloud Environments such as Hadoop Cluster File System and MapReduce programming model. The parallelized MapReduce involves many data analysis algorithms for effective analysis of data. HDFS is Hadoop Distributed File System which can process the operations in distributed environment and it holds huge data sets such as terabytes or even petabytes and also provides extensive level of throughput access.

Sanjeev Dhawan et al. [24] presented Apache Pig with MapReduce which is used for getting good results with better analysis techniques. For getting results, the MapReduce job was generated with Hive framework in the big database. The performed analysis by MapReduce machine is successfully shown as final results. The performance of both was nearly same.

G Geethakumari et al. [25] proposed data security model for investigating the impact of Big Data techniques. The authors also proposed for developing accurate secure analytical techniques to analyze about security threats.

III. BIG DATA ANALYTICS

A big problem is created with Big Data because large size, variety and data speed. So, it is analyzed properly by using Big Data Analytic Tools. Big Data Analytics is the specific mechanical process to analyze data for bringing out hidden content to outside world that can be extracted to take better decisions. The potential use cases for Big Data Analytics are shown in Figure 1. It describes the relation between the different kinds of data. The data is represented in structured form, unstructured and semi- structured form with velocity of data from batch system to real time system. Big

Data Analytics presents the different analysis with various levels.

The data analysts and scientists are used Big Data Analytics to analyze large volumes of data. Consider an enterprise; Hundreds of millions of data combinations in multiple data stores and different formats and billions of rows of data is maintained by an enterprise [26]. The value of Big Data Analytics by generating graph between time and cumulative cash flow is depicted in Figure 2. The existing analytical techniques are required more number of hours or days to analyze and process data as compared with Big Data Analytics.

If the information is analyzed and processed in the right time then it can have value otherwise no useful, so its value is not returned back at proper cost.

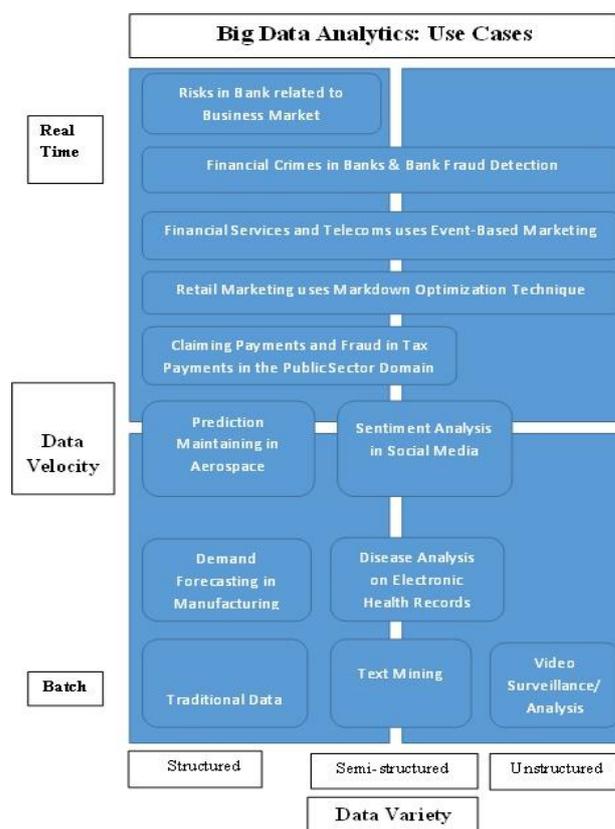
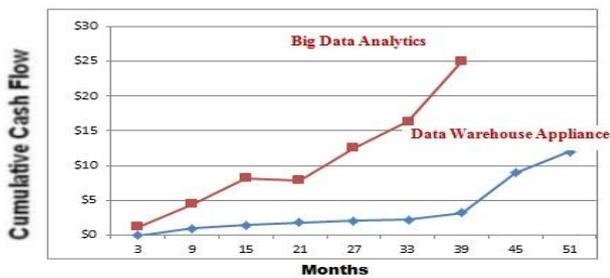


Figure.1 Use Cases for Big Data Analytics

The most difficult and challenging business problems can be tackled easily with new advancement in computing technology. High-performance analytics are used to process the relevant data in the simple way [8]. The more reliable and suitable innovation techniques are extracted to take best possible decision by using high-performance predictive analytics. The organizations, industries and business enterprises are finding for reliable analytics to meet the requirements of big data [10].

Hadoop Framework for Big Data Analytics to Analyze Attacks on Business Data



- Projected cash flow data actual business case analysis and real world experiences
- Breakeven 4 mos. vs.8
- 3-year Project Value \$25M vs. \$12M

Figure 2 Values of Big Data Analytics

Big Data Analytics can help to manage the diversified data. It can also help in the generalizing of new queries from different tests, formulating new hypotheses, exploring and discovering the new processed concepts, and taking decisions.. The big data analytics are applied for either new data or mixed data. Big data Analytics is used Hadoop tool to process the unstructured data. Hadoop is already proved its advantages of use and cost [11]. Cloudera, Hortonworks, and MapReduce can work on Hadoop framework with high-scale storage capacity. Big Data analysis is main requirement for today's businesses world. Hadoop is a high-end technology to analyse on traditional databases [18]. Today, Big Data Analytics are used by only 8 percent of large global organizations to identify patterns. Some attacks are done on payment systems by compromising both customers' financial information and status of enterprise. By 2016, 25 percent of companies can use Big Data Analytics.

IV. HADOOP

Hadoop was built on java platform which is open source tool. It is a software framework which is used to store and process Big Data using Big Data analytics [9]. The main aim of Hadoop is to improve the data processing performance on clusters.

- Hadoop is performed processing of large data sets easily and faster. It encompasses various concepts such as HDFS, Map-Reduce, HBASE, PIG, HIVE, SQOOP and ZOOKEEPER [13].
- Hadoop is generated high value by processing with huge size, more speed and high different types of data [18].

Figure 3 describes that how Hadoop took the different type of data, processed and analysed the entire data and implement action immediately. Hadoop uses NoSQL or MongoDB languages to process the data.

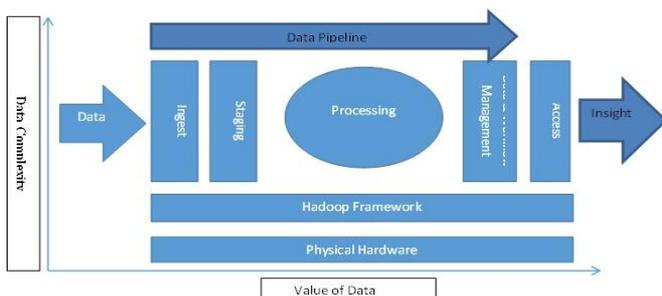


Figure 3 Hadoop converts Big Data into high value

Hadoop Framework was designed to work in parallel on large data sets and connecting with more number of commodity computers in efficient manner [11]. The cost of 1000-CPU (processor) machines is very huge amount of money. Instead of using one thousand individual CPUs or 250 quad-core machines, Hadoop is tied simple and less cost machines collectively as a single cost-effective computer cluster [8]. In Hadoop cluster, data can be loaded and distributed to all the cluster nodes as shown in Figure 4. The Hadoop Distributed File System (HDFS) can be decomposed the large data files into small units and these can be monitored with other nodes in the cluster [9]. In the case of system failures, the data chunks are re-replicated by an active monitoring system. The replicated and distributed data chunks formed a single namespace across the number of machines, so these are universally accessible [17].

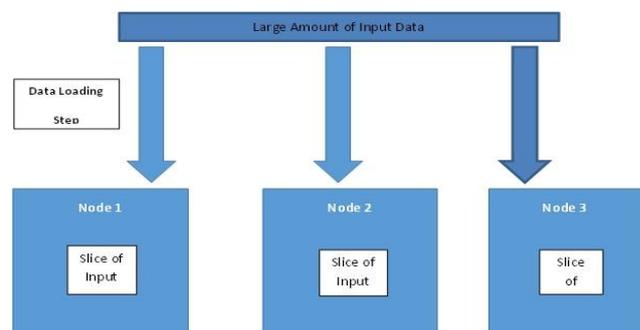


Figure 4 Hadoop Cluster

Hadoop programming framework is having the cluster data as conceptually record-oriented where the distributed file system stores files. Every machine is performed an operation on a subset of the data. Hadoop produces results in high performance because of achieving high data locality where Hadoop fits with the moving computation strategy [22]. Hadoop communicates with individual processes in fewer manners. Isolated task processes each record separately particularly within the domain. MapReduce is a programming model to process records separately as isolated tasks called Mappers [19]. The output of Mappers is transferred to Reducers, to merge the different mappers together which are shown in Figure 5.

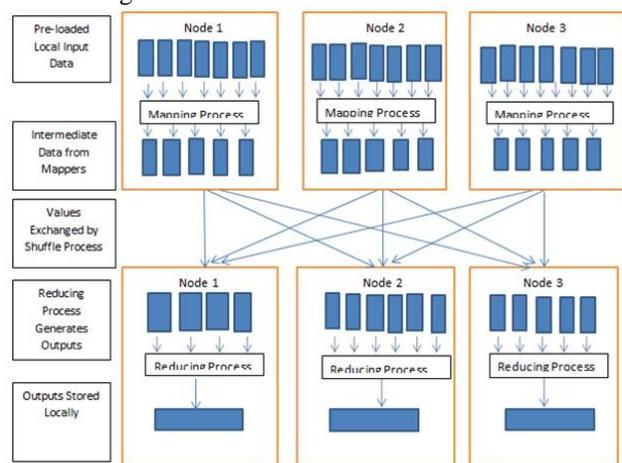


Figure 5 Hadoop Distributed File System maps records with different nodes

The individual servers in a cluster are allowed by Hadoop Distributed File System without stopping the process which data duplication across the cluster to get failure [14]. The HDFS stores data without any limit, any form of data. The relational databases must be structured the data with schema definition before storing. The main responsibility of developer's code using HDFS is to create the value and sense of the data [19].

V. BIG DATA SECURITY ANALYTICS

There are many opportunities for big data security analytics because of:

1. Continuous detection to identify the problems. Existing security tools may not analyzed the global cyber-attack which includes advanced patterns of virus, malware, stealthy attack techniques [25]. Even though, the enterprises bought advanced anti-virus and malware tools, the Big Data Security Analytics with real time tools are enhanced to defense the threats and malwares [26].
2. The cost of multi-core 64-bit Intel servers with 10 GBPS network interfaces is cheaper than old servers and these are faster and size is too small comparatively than the old systems. The computing power of these servers is more for stream and batch processing with massive data. By customizing open source tools, Security vendors are increasing development cycles for security analytics purposes. The big data analytical tools are MapReduce, and Mahout, Cassandra, Hadoop [20]. This can help for innovate advanced security methods to protect systems from different threats.
3. The Enterprises are required security alerts from vendors. The vendors like ISC8, Narus, Leidos, and Palantir, 21CT, Hexis Cyber Solutions are turned up into private sector [16]. The hackers are continued to hack information with timeliness property. The hackers could not success their targets forever when it uses effectively [20]. Before two years, the hackers had to target mainly to theft for information or money by conducting extensive cyber espionage.

Now a days the hackers are too intelligent to find advanced secure principles and fraud preventive techniques. These are configured in the enterprises. According to Gartner words, the existing monitoring systems may not produce correct results in reducing false alarms. The top highest priority alerts are correlated to identify fraud related patterns and abuse. By adopting these models and rules, the speed is increased for getting their responses [12].

VI. ISSUES

There are many Big Data concepts for research in the world. Even though, researchers are doing a lot of research work on Big Data to solve the problems, but there are many concepts still in pending for research. The researchers have to increase security system for improving the software capability to identify updated malicious programs and threats and design advanced threat detection systems [19]. Reliable and quality security system is required to prevent threatening for business organizations. So, researchers have to design such kind of security system. By using machine learning techniques, the researchers plan for collecting data, pre-treatment, and integration of data [23]. For providing security and

implementing preventive measures, the researchers can use the results to analyse accordingly. Researchers have to use Big Data Security Analytics with Hadoop to design high quality security system to prevent threats [24]. Packet pig and Mahout are networking monitoring tools which are used for enhancing the secure levels. The researchers have planning to design advanced security threat prevention system, threat detection system and analysed effectively using Hadoop cluster technology. The system efficiency is evaluated to design threat preventive measuring system using security analysis [21].

VII. CONCLUSION

In our research, we have proposed necessity of Analytics for Big Data for analyzing business organization and enterprise data. The security framework for Big Data Security Analytics using Hadoop structure to deal the malicious threats and non-malicious threats was presented. We have described the characteristics of Big Data for huge sets of business enterprise data. The high efficiency analytics model and Hadoop can provide good opportunities significantly to meet the business needs of an enterprise if it does not for making the strategic decisions.

FUTURE PLANS

Providing security for enterprise and organizational data is more challenging task in the implementation of security policy formulation and mechanisms.

By using machine learning techniques, we can implement for data collection, pre-treatment, integrating data and predicting data. In future of my research, we can develop advanced security alerts before occurring the target on specific data and so every employee can get the filtered events and summarization of view.

CONFLICT OF INTEREST:

There is no conflict of interest from authors.

ACKNOWLEDGMENT:

I thank to all who supported me to write this paper successfully. I am thankful to all the authors of research papers which I was referred to complete my research work. I specially thank to my research supervisor for his full support and good advices to do this work.

REFERENCES

1. Mr. Mohan Shivaji Kamble, Prof. Vaishali Londhe, "3-D Face Image identification from video streaming using Map Reduce(Hadoop)", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 03 | Mar-2018, e-ISSN: 2395-0056.
2. K. Tamilselvi1, V. Sumithra2, Mrs. K. Dhanapriyadharsini3, "Big Data Analytics Using Hadoop Technology", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 01 | Jan-2018, e-ISSN: 2395-0056.
3. Shaili, Durgesh Srivastava, Deepak Sinwar, "Statistical Analysis of Big Data Using Hadoop: A Review", IAETSD Journal For Advanced Research In Applied Sciences, Volume 5, Issue 1, Jan/2018, ISSN NO: 2394-8442.

Hadoop Framework for Big Data Analytics to Analyze Attacks on Business Data

4. Suruchi Padhy, Dr. Shashi Kumar D R, “**Big Data Analysis Using Apache Hadoop**”, International Journal of Advance Research, Ideas and Innovations in Technology. Volume 4, Issue 1, 2018, pp- 225-227, ISSN: 2454-132X,
5. K.Venkatesh, Md.Ahamed, “**Privacy Preserving Enriched Map Reduce for Hadoop Based Big Data Applications**”, International Journal of Recent Trends in Engineering and Research, 2017, ISSN-2455-1457.
6. Nishant Rajput, Nikhil Ganage, Jeet Bhavesh Thakur, “**Review Paper on Hadoop and Map Reduce**”, International Journal of Research in Engineering and Technology, Volume: 06 Issue: 09 | Sep-2017, e ISSN: 2319-1163.
7. Vaishnavi Chillal, Nikita Dalvi, Bhagyashree Nagapure, Shamla Naidu, Prof. Rajesh Bharati, “**Big Data Service Architecture for E-portal**”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 4, April 2017, ISSN (Online) 2278-1021.
8. Prof. Girija Chiddarwar, Sanket Chhajed, Sameer Deshmukh, Pranoti Dongre, Bharti Nile, “**Big Data Analysis of Fetched Logs using Hadoop Framework**”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 3, March 2017, ISSN (Online) 2278-1021.
9. Goddilla NagarjunaReddy, M.V.Jagannatha Reddy, “**Big Data Processing Using Hadoop in Retail Domain**”, International Journal Of Engineering And Computer Science, Volume 5 Issue 09 September, 2016 Page No.18175-18179, SSN: 2319-7242.
10. L. Greeshma, G. Pradeepini, “**Big Data Analytics with Apache Hadoop MapReduce Framework**”, Indian Journal of Science and Technology, Vol 9(26), July 2016, ISSN (Online): 0974-5645.
11. Saloni Minochal, Jitender Kumar, S Hari Singh, Seema Bawa4, “**HadoopWeb: MapReduce Platform for Big Data Analysis**”, International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 07 | July-2016, e-ISSN: 2395-0056.
12. Dr. Urmila R. Pol, “**Big Data Analysis Using Hadoop Mapreduce**”, American Journal of Engineering Research (AJER), Volume-5, Issue-6, 2016, pp-146-151, e-ISSN: 2320-0847.
13. Dr. Urmila R. Pol, “**Big Data Analysis: Comparison of Hadoop MapReduce, Pig and Hive**”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 6, June 2016, ISSN (Online): 2319-8753.
14. M. Dhavapriya, N. Yasodha, “**Big Data Analytics: Challenges and Solutions Using Hadoop, Map Reduce and Big Table**”, International Journal of Computer Science Trends and Technology (IJCT) – Volume 4 Issue 1, Jan - Feb 2016, ISSN: 2347-8578.
15. Amol Bansod, “**Efficient Big Data Analysis with Apache Spark in HDFS**”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-4 Issue-6, August 2015.
16. Joelson A. dos Santos, Talat Iqbal Syed, Murilo C. Naldi, Ricardo J. G. B. Campello, Jorg Sander, “**Hierarchical Density-Based Clustering using MapReduce**”, IEEE. Translations, VOL. 14, NO. 8, AUGUST 2015, ISSN: 2332-7790.
17. Rotsnarani Sethy, Mrutyunjaya Panda, “**Big Data Analysis using Hadoop: A Survey**”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 7, July 2015, ISSN: 2277 128X.
18. Rahul Beakta, “**Big Data and Hadoop: A Review Paper**”, RIEECE -2015, Volume 2, Spl. Issue 2 (2015), e-ISSN: 1694-2329.
19. Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar, “**A Review Paper on Big Data Hadoop**”, International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014 1 ISSN 2250-3153.
20. Bhawna Gupta, Dr. Kiran Jyoti, “**Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data**”, International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 3867-3870, ISSN: 0975-9646.
21. Dr. Siddaraju, Sowmya C L, Rashmi K, Rahul M, “**Efficient Analysis of Big Data Using Map Reduce Framework**”, International Journal of Recent Development in Engineering and Technology, Volume 2, Issue 6, June 2014, ISSN 2347-6435.
22. Vidyullatha Pellakuri, Dr.D. Rajeswara Rao, “**Hadoop Mapreduce Framework in Big Data Analytics**”, International Journal of Computer Trends and Technology (IJCT) – volume 8 number 3–Feb 2014, ISSN: 2231-2803.
23. Rabi Prasad Padhy, “**Big Data Processing with Hadoop-MapReduce in Cloud Systems**”, International Journal of Cloud Computing and Services Science (IJ-CLOSER), Vol.2, No.1, February 2013, pp. 16~27 ISSN: 2089-3337.
24. Sanjeev Dhawan, Sanjay Rathee, “**Big Data Analytics using Hadoop Components like Pig and Hive**”, American International Journal of Research in Science, Technology, Engineering & Mathematics(AIJRSTEM), 2013, ISSN (Online): 2328-3580.
25. G Geethakumari, Agrima Srivatsava, “**Big Data Analysis for Implementation of Enterprise Data Security**”, International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, No.4, August 2012, ISSN: 2249-9555.
26. Siva Sankara Reddy Donthi Reddy, Udaya Kumar Ramanadham, “**Big Data Analytics for Healthcare Organization, BDA Process, Benefits and Challenges of BDA: A Review**”, Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 4, 189-196 (2017).

AUTHORS PROFILE



Siva Sankara Reddy Donthi Reddy received his ME degree in Computer Science and Engineering from Sathyabama Institute of Technology and Science, Deemed to be University, Chennai in 2007 and he is presently doing his PhD degree in Computer Science and Engineering from Bharath Institute of Higher Education and Research, Chennai. His research interests include Big Data, Cloud Computing and Cryptography and Network Security.



Udaya Kumar Ramanadham received his PhD degree in Computer Applications from Bharath Institute of Higher Education and Research, Deemed to be University, Chennai, India. He is currently Professor in the Department of Information Technology at Bharath Institute of Higher Education and Research, Deemed to be University, Chennai. His research interests lie in the areas of Data Mining, Big Data, Network Security and Computer Networks.