Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 3, March 2016, pg.54 – 58

ANALYSIS OF VARIOUS BIG DATA TECHNIQUES FOR SECURITY

Suriya Begum[1], Kavya Sulegaon[2]

[1]Prof., Department of Computer Science and Engineering, New Horizon College of Engineering, VTU, India

[2]Student, Department of Computer Science and Engineering, New Horizon College of Engineering, VTU, India

suriyabegumnotes@gmail.com

Abstract – The data is growing day by day to a larger extent. Recently, Big Data Analytics has become a hot topic in academics, industry and everywhere. Big Data Analytics is the process of examining large amounts of data (big data) to discover hidden patterns or unknown correlations. A key part of big data analytics is the need to collect, maintain and analyse enormous amounts of data efficiently. Due to increase in number of sophisticated targeted threats and rapid growth in data, the analysis of data becomes too difficult. Today's attacks are prepared by advanced technologies are not detected until the damage has been occurred. Big Data Security Analytics is important to mitigate the security threats to secure the data more efficiently. In Big Data Analytics, Data Security is a challenging task to implement and calls for strong support in terms of security policy formulation and mechanisms. In this paper we have discussed the analysis of the Big Data Analytics concepts and some existing techniques and tools, like Hadoop, for data Security.

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

I. INTRODUCTION

Big data refers to data sets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. With the evolution of technology and the increased multitudes of data flowing in and out of organizations daily, there has become a need for faster and more efficient ways of analyzing such data [15]. Big Data Analytics (BDA) can also enable the construction of predictive models for customer behavior and purchase patterns, therefore raising overall profitability [13]. Big data can create transparency, and make relevant data more easily accessible to

© 2016, IJCSMC All Rights Reserved

stakeholders in a timely manner. Big Data Analytics holds much potential for customer intelligence, and can highly benefit industries such as retail, banking, and telecommunications [17]. In Big Data Analytics, Data Security is a challenging task to implement and needs strong support in terms of security policy formulation, techniques, tools and mechanisms. Hadoop is one of the tools and is a highly scalable storage platform, because it can store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel [2,3,8,18].

II. EXISTING BIG DATA ANALYTICS TECHNIQUES AND TOOLS FOR SECURITY

A. **Nada Elgendy [1]** analyses some of the different analytics methods and tools which can be applied to Big Data (BD), as well as the opportunities provided by the application of Big Data Analytics (BDA) in various decision domains. Big Data tools, techniques, and governance processes can increase the prevention and recovery of fraudulent transactions by dramatically increasing the speed of identification and detection of compliance patterns within all available data sets. It also discussed about some of the different advanced data analytics techniques. BD,as well as its characteristics and importance has been discussed in [7]. BD is largely untagged file-based and unstructured data, about which little is known .This means not only that large quantities of potentially useful data is getting lost[17].

B. **Bhawna Gupta [2]** proposes the use of BDA for analysing the enterprise data. The main focus is to gather the unstructured data from all the terminals, processed the data to convert into structured form so that accessing of the data would be easier. BDA describes the simple algorithm for large amount of data without compromising performance. Hadoop is one of the tools which are aimed to improve the performance of data processing. In this approach they are managing the Big Data characteristics of large volumes of enterprise data. If enterprise has an unmet business need for strategic decision making with a high degree of processing, a Revolution Analytics and Hadoop combination offers significant opportunity to gain advantage [8]. Hadoop is a highly scalable storage platform, because it can store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel.

C. **Weiyi Shang et. al .[3]** describes a first step in assisting developers of big data applications BDA Apps for cloud deployments. It proposes a lightweight approachfor uncovering differences between pseudo and large-scale cloud deployments. Using injected deployment faults; they have shown that their approach is not only significantly reduces the deployment verification effort, but also provides very few false positives when identifying deployment failures. It proposes an approach for verifying the runtime execution of BDA Apps after deployment. The approach abstracts the platform's execution logs from both the small and large scale cloud deployments, groups the related abstracted log lines into execution sequences for both deployments, then examines and reports the differences between the two sets of execution sequences. The Authors specifies that the larger data and more complex environments lead to unexpected executions of the underlying platform. Such unexpected executions and their context cannot be easily uncovered by traditional approaches. In this paper, they propose an approach to uncover the different behaviour of the underlying platforms for BDA Apps between runs with small testing data and large real-life data in a cloud environment. To evaluate the approach, they

© 2016, IJCSMC All Rights Reserved

have performed a case study on Hadoop, a widely used platform, with three BDA Apps [9,16]. BDA Apps are a new category of software applications that leverage large-scale data, which is typically too large to fit in memory or even on one hard drive, to uncover actionable knowledge using large scale parallel-processing infrastructures.

D. **Ulla Gain1 [4]** develops BD and symbolizes the aspiration to build platforms and tools to ingest, store and analyze data that can be voluminous, diverse, and possibly fast changing.This strategy is partly descriptive and partly improving. Through launching the term data-milling the Authors try to improve understanding of the phenomenon of BD, as well as, possibilities of data analytics. Launched the term data-milling to represent the searching of the information nuggets from the heterogeneous data. To justify the launched term data-milling, they made the literature review in which they searched the definitions of BDA. Their study shows that BDA is verbosely explained. They used only four statements from 19 to crystallize BDA. The literature review of BDA gave the description of current status of the phenomenon BD. The launched term data-milling improves the understanding of the phenomenon BD, as well as, possibilities of data analytics [10].There exist large amounts of heterogeneous digital data.This phenomenon is called BD which will be examined. The examination of BD has been launched as BDA.

E. **Alexander Ginsburg et.al [5]** describes the term BD to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies. BD is changing security analytics by providing new tools and opportunities for leveraging large quantities of structured and unstructured data. The Authors specifies the differences between traditional analytics and BDA, and briefly discusses tools used in BDA. They also proposes a series of open questions about the role of BD in security analytics. Big Data technologies can be divided into two groups: 1) Batch Processing, which are analytics on data at rest, and 2) Stream Processing, which are analytics on data in motion. The Authors proposes security to BD by resolving the BDA issues, such as , 1) Data Provenance , which provides the Authenticity and Integrity of data used for analytics. 2) Privacy which enhances a method for regulatory incentives and technical mechanisms to minimize the amount of inferences that BD users can make. 3) Securing Big Data stores ,which focuses on using BD for security, but the other side of the coin is the security of BD.4) Human-computer interaction, which mentions that the BD facilitates the analysis of diverse sources of data. Compared to the technical mechanisms developed for efficient computation and storage, the human-computer interaction with BD has received less attention and this is an area that needs to grow [11]. The approach is to treat products and services as parts of complex systems that consist of both social and technological components. Human-computer interfaces are an integral part of the functioning of these systems.

F. **Jainendra Singh [6]** discusses about Machine Learning (ML) techniques which have found widespread applications and implementations in security issues. Machine Learning algorithms are used in very diverse contexts: 1) to recognize handwritten text, 2) to extract information from images, 3) to build automatic language translation systems, 4) to predict the behavior of customers in an online shop, 5) to find genes that might be related to a particular disease, and so on. This approach focuses on the

© 2016, IJCSMC All Rights Reserved

development of fast and efficient algorithms for real-time processing of data as a main goal to deliver accurate predictions of various kinds. ML techniques can solve the above mentioned applications using a set of generic methods that differ from more traditional statistical techniques. It specifies that the advancement in ML, provides new challenges and solutions to the security problems encountered in applications, technologies and theories [12]. ML is ideal for exploiting the opportunities hidden in BD. It delivers on the promise of extracting value from big and disparate data sources with far less reliance on human direction. It is data driven and runs at machine scale [18, 19].

III. CONCLUSION

As the data is becoming bigger and bigger, there is a need to store this data in an efficient manner. I this paper, we have examined the innovative topic of big Data, which has recently gained lots of interest due to its applications. An analysis has been done on BDA, in order to provide an insight on the BDA concepts. Data Security is a challenging task to implement and calls for strong support in terms of security policy formulation and mechanisms. We plan to take up data collection, pre-treatment, integration, map reduce and prediction using Machine Learning techniques. In future, we are planning to develop security alerts, which will provide employees with the ability to view the activity. Events will be filtered down and summarized view will be available to each individual employee.

REFERENCES

[1] Nada Elgendy, Ahmed Elragal ,"Big Data Analytics: A Literature Review Paper", ICDM, LNAI 8557, pp. 214–227, 2014

[2] Bhawna Gupta , Dr. KiranJyoti ,"*Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data*", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3867-38702014

[3] Weiyi Shang , Zhen Ming Jiang , Hadi Hemmati , Bram Adams , Ahmed E. Hassan , Patrick Martin "*Assisting Developers of Big Data Analytics Applications When Deploying on Hadoop Clouds*",IEEE 978-1-4673-3076-3/13 IEEE,2013

[4] Ulla Gain1 ,VirpiHotti ,"*Big Data Analytics for Professionals, Data-milling for Laypeople* ",International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 4, Number 1 (2014), pp. 33-402013

[5] Alexander Ginsburg, Luciano JR Santos, KendallScoboria, Evan Scoboria, John Yeoh, "*Big Data Analytics for Security Intelligence"*, 2014

[6] Jainendra Singh , "*Big Data Analytic and Mining with Machine Learning Algorithm*", World Journal of Computer Application and Technology 1(2): 51 -57, DOI: 10.13189/wjcat.2013.010205,2014

[7] www.forensicrisk.com/big-data-analytics-and-fraud-prevention

[8] www.itproportal.com/big-data-5-major-advantages-of-hadoop

[9] http://web.cs.ucla.edu/~miryung/teaching/EE379K-Spring2014/Papers/Paper%207.pdf

[10] http://www.techrepublic.com/resource-library/whitepapers/big-data-analytics-for-professionals-data-milling-for-laypeople

© 2016, IJCSMC All Rights Reserved

[11] http://researcher.watson.ibm.com/researcher/view_group.php?id=4

[12] http://www.skytree.net/machine-learning/why-do-machine-learning-big-data

[13] https://en.wikipedia.org/wiki/Big_data

[14] https://mobilesecuritywiki.com/

[15] https://en.wikipedia.org/wiki/Big_data

[16] Vignesh Prajapati , " Big Data Analytics with R and Hadoop ", Kindle Edition.

[17] Seema Acharya and Subhashini Chellappan ," Big Data and Analytics " , Apr 2015

[18] Tom M. Mitchell , " Machine Learning ", May 2013.

[19] ETHEM ALPAYDIN , "Introduction to Machine Learning ",3rd Edition, Hardcover – 2015

Other Publications :

[20] Suriya Begum ,Prashanth ―Review of Load Balancing in Cloud Computing‖ ,IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013 ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814

[21] Suriya Begum,Prashanth ―Investigational Study of 7 Effective Schemes of Load Balancing in Cloud Computing ―,IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 1, November 2013 ,ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784

[22] Suriya Begum, Dr. Prashanth C.S.R,―Mathematical Modelling of Joint Routing and Scheduling for an Effective Load Balancing in Cloud‖ in International Journal of Computer Applications (0975 –8887) Volume 104 –No.4, October 2014.

© 2016, IJCSMC All Rights Reserved