



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: IV Month of publication: April 2018

DOI: <http://doi.org/10.22214/ijraset.2018.4264>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

A Study on Log Parser Analysis and Error Detection using Big Data

Dileep V¹, Dr. S. PathurNisha²

¹U. G. Scholar, Department of Computer Science and Engineering, Nehru Institute of Technology, Coimbatore.

²Professor, Department of Computer Science and Engineering, Nehru Institute of Technology, Coimbatore.

Abstract: Log examination could be a method for mechanically understanding the important examples from heterogeneous Log information. Each movement occurring in an application or gadget is recorded in a log document. Valuable information is stored on log data which can be extracted and stored into big data platforms. Prediction and classification can be performed over the log data.

Keywords: Log analysis, prediction, big data and hadoop.

I. INTRODUCTION

Log files provide valuable information about the functioning and performance of applications and devices. These files are used by the developer to monitor, debug and troubleshoot the errors that may have occurred in the application. Manual processing of log data requires a huge amount of time, and hence it can be a tedious task. The structure of error logs vary from one application to another. Since volume, velocity and variety being dealt here, big data using hadoop is used. Analytics involve meaningful and understanding patterns from various log files.

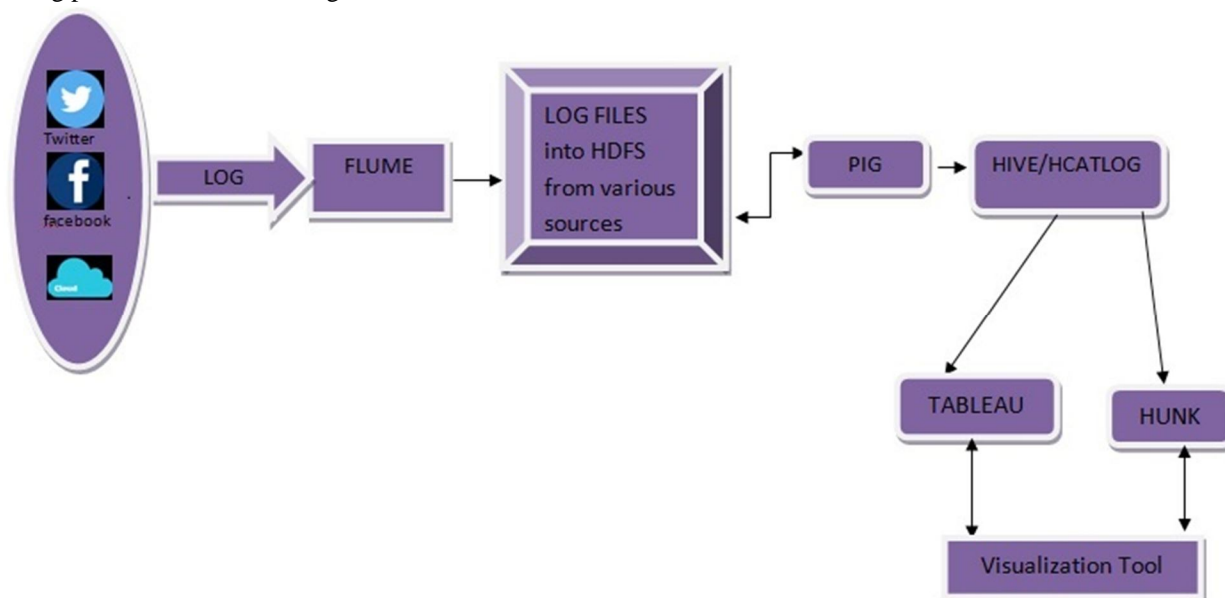


Fig 1. System Architecture

Error log analytics deals regarding the conversion of knowledge from semi structured to constant structured format, fixed analytics are going to be performed over it. Business analytics is employed to predict and forecast the longer term standing of the applying supported current .Proactive, measures can be taken rather than reactive, measures in order to efficient maintainability of the application and devices.

II. LITERATURE REVIEW REPRESENTATION

Amit Dipchandi and Dr.Girish S. Kalkar (2014) have proposed that the availability of the data of web accessed is in human readable form generated by computer referred to as web log, provided by online sources, it make that data into day to day life of individual as

well as for business operations for more dynamism and bring it closer to real time for the web administrator about what is happening with the web. With the help of such web log data helping the business organization before having to wait a week, or even a month for data through which those people will be able to mine data and perform predictive analysis from multiple access made daily as well as in regular manner from users around the world. Data Mining is used for finding expected patterns from that large set of log data using Web Mining. When used together, predictive analytics and data mining can make the future prediction more efficiently with respect to web access.

Prajakata Dange and Dr. Deven Shah (2015) have inspected that a rapid development of Web, there is enhancement in information sharing through social networking and increasing business adoption of the Web as a means of doing business and delivering service and thereby increases website attack often. Hackers either seek to compromise the corporate network. As a result, there is need to pay increased attention to the security of the web applications. The majority of web application attacks occur through cross-site scripting (XSS) and SQL injection attacks. There are some protection mechanisms like firewalls were not designed to protect web application and thus may not be provide proper solution to attack problem. The proposed system emphasis on SQL injection and Cross site scripting .i.e. to detect the most critical web application security flaws using log file. The system uses Apache log for this. Web log file can be helpful for detailed analysis of user action. Log file is automatically created and maintained by a server of activity performed by it. This information may be combined into one file, or separated into distinct logs, like an access log, error log, or referrer log. The majority of attacks can be recognized and acted upon to prevent further exploitation and to reduce risk. Proposed system parse a log record and get compare with the help of regular expression and identify the malicious record.

M. Vithaya and Dr. S. Sugana (2016) discussed that Web is an important part of organization. Every organization generated huge amount of data from various source. Web mining is the process of discovering the knowledge from the web data. The log files are maintained by the web server. Analyzing web log files has become an important task for E-commerce companies to predict their customer behavior and to improve their business. E-commerce website can generate tens of Petabytes of data in their web log files. So, the large volume of data is called big data. Big data is something so huge and complex that is impossible for handling through traditional system and traditional tools. The analysis of log files is used for learning the user behavior. The analysis of such large web log files are be worked upon by using traditional SQL does not like queries nor can the relational database management system (RDBMS) be used for storage and analysis. So, need parallel processing and reliable data storage system for this huge and complex data. The Hadoop framework provides reliable storage by Hadoop Distributed File System and parallel processing system for large database using Map reduce programming model. This mechanism helps to process log data in parallel using all the machines in the Hadoop cluster and computes results efficiently.

Neha Goel and C.K. Jha(2013) have discussed that Internet is acting as a major source of data. As the number of web pages continues to grow the web provides the data miners with just the right ingredients for extracting information. In order to cater to this growing need a special term called Web mining was coined. Web mining makes use of data mining techniques and deciphers potentially useful information from web data. Web Usage mining deals with understanding the behavior of users by making use of Web Access Logs that are generated on the server while the user is accessing the website. A Web access log contains of varied entries just like the name of the user, his information science address, range of bytes transferred timestamp etc. An assortment of Log Analyzer devices exist which help in breaking down different things like clients navigational example, the piece of the site the clients are for the most part inspired by and so on this paper influences utilization of such log to instrument apparatus alluded to as blog learned for finding out the conduct of clients United Nations organization get to relate pseudoscience site.

Pushkar Gavandi, Bhavika Gori, Smruti Ingawale, and Seema Yadav (2016) have proposed that Big Data is an emerging growing dataset beyond the ability of a traditional database tool. Hadoop rides the massive information wherever the large amount of data is processed victimization cluster of trade goods hardware. A web server log record is a content document that is composed as action is produced by the web server. Log records gather an assortment of information about data solicitations to your web server. Server logs go about as a guest sign-in sheet. Server log documents can give data about what pages get the most and the minimum movement. What pages that your guests see and the programs and working frameworks used to get to your site. The web server log handling has splendid, dynamic degree in the field of data innovation. The web server log preparing can be so upgraded and extended that it can be utilized as a part of different spectra's and fields which are dealing with huge measure of information on everyday schedule. It is dependable, quick and versatile approach for dealing with expansive quantities of logs and to change log information into measurable information and produce reports as needs be.

Shlomi Dolev, Ehud Gudes, Shantanu, and Ido Singer (2017) have presented a survey that Hadoop and Spark square measure wide used distributed process frameworks for large-scale processing in an economical. Hadoop and Spark are widely used distributed processing frameworks for large-scale data processing in an efficient and fault-tolerant manner on private or public clouds. These

enormous information handling frameworks are widely utilized by numerous enterprises, e.g., Google, Facebook, and Amazon, for taking care of an expansive class of issues, e.g., seek, grouping, log examination, varying kinds of be a piece of tasks, grid activity, design coordinating, and interpersonal organization investigation. Be that as it may, all these well known frameworks have a noteworthy downside as far as privately dispersed calculations, which avert them in actualizing geologically conveyed information preparing. The expanding measure of topographically dispersed gigantic information is pushing businesses and area to reexamine this huge information strategy framework. The novel structures, which will be past best in class designs and innovations associated with the present framework, are relied upon to process geologically conveyed information at their areas without moving whole crude datasets to a solitary area. Savitha K, Vijay MS (2017) have investigated the role of Log analysis helps to improve the business strategies as well as to generate statistical reports. Hadoop map reduce based log file analysis tool will provide as graphical reports showing hits for web pages, users page view activities in which part of website users are interested, traffic, attack etc.

III. CONCLUSION

The information obtained with these experiments show the effectiveness of the web usage mining application for predictive analytics in the businesses, not only in reducing considerably the size of Web log files but also in grouping Web requests into a number of user which can encode the user browsing behavior in a significant manner. After having information about access the web site architecture can be improved thus helping the company to take future decision by doing predictions. It all can solve the queries like how to describe the preferences of users on the basis of their navigational access. Several measures or heuristics can be applied to obtain the degree of interest for a Web resource. A possibility is to consider the degree of interest for predictive analysis to a resource as strictly related to the frequency of accesses to that resource (number of accesses to that resource or total number of accesses during the session) and to the time the user spends on the same one. Analysis of this valuable information will help companies to develop plans that are more effective and impactful, easy access to website, in between company communication and productive skills through predictive analytics when used with web usage mining.

REFERENCES

- [1] AmitDipchandi and Dr.Girish S. Kalkar (May 2014), "Use of Log Data For Predictive Analytics Through Data Mining" Current Trends in Technology and Science, Vol No 3, Issue no 3.
- [2] B. Chandamouli, J. Goldstein, D. Songyun (2012), "Temporal Analytics on Big Data for Web Advertising", IEEE 28th International Conference on Data Engineering (ICDE), Volpp.
- [3] J. Dean and S. Ghemawat(2008) "MapReduce: simplified data processing on large clusters", Communications of the ACM 51(1):Page no107-113,
- [4] L.K.J.Grace,V.Maheswari,D.Nagamalai,(2011),"Analysis of Web Logs and Web User in Web Mining", International Journal of Network Security and its applications(IJNSA). Vol 3, Issue no 1.
- [5] Maryam Jafari, ShahramJamali (2013), " Discovering Users Access Patterns for web Usage Mining from Web Log Files" Journal of Advanced in Computer Research vol 4, issue 3,
- [6] M.Vithaya and Dr.S.Sugana (2016), "Analysis Of Web Logs Using Big Data Tools" International Journal Of Advanced Research trends In Engineering and Technology, Vol 3, Special Issue 20.
- [7] MsShashiSahu, LeenaSahu (2015)" A Survey on Frequent Web Mining with Improving Data Quality of Log Cleaner" International Journal of Advanced Research in Computer Engineering & Technology, vol 4, Issue no.3.
- [8] NehaGoel and C.K. Jha(2013), " Analyzing Users Behavior from Web Access Logs using Automated Log Analyzer Tool"International Journal of Computer Applications (0975 – 8887) Volume 62– No.2.
- [9] PrajakataDange and Dr.Deven Shah (2015), "Web Log Analysis For Security Compliance Using Big Data" International Journal Of Advanced Research in Computer Science and Software Engineering, Vol.5, Issue no.3.
- [10] PushkarGavandi, BhavikaGori, SmrutiIngawale, andSeemaYadav(2016),"Web Server Log Processing using Hadoop" International Journal for Research in Engineering Application & Management (IJREAM)Vol-01, Issue 10.
- [11] Savitha K, Vijay MS (2014),"Mining of web server logs in a distributed cluster using Big data Technologies" IJACSA Vol.5.
- [12] S. Sathya, M. Victor Jose(2011), "Application of HadoopMapReduce Technique to Virtual Database System Design", International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), Volpp,Page no 892-896, 2011.
- [13] SholmiDolev, Patricia Florissi, Ehud Gudes, Shantanu Sharma and Ido Singer (2017), "A Survey on Geographically Distributed Big data processing Using Mapreduce" IEEE Transactions on Big Data ,Vol :PP ,Issue no.99.
- [14] S.SiddharthAdhikari, DeveshSaraf, Mahesh Revanwar and Nikhil Ankam (2014), "Analysis Of Log Data and Stastics Report Generation Using Hadoop" International Journal of Innovative Research in Computer and Communication Engineering, Vol 2, Issue no. 4.
- [15] Xindong Wu AT,(2014), "Data Mining with Big Data" IEEE Transcations on Knowledge and data Engineering".Vol 26,Issue no.1.
- [16] www.google.com



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)