International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

Web Server Log Analyzer

[1.] Nisar Nadaf, [2.] Pramod Pathak
[1,2.] Student, BE-Computer Engg
Smt.Kashibai Navale College of Engg,Pune, India

***Abstract*— Large amount of useful operations data and warnings of failures are available in the server logs. The challenge is that such information gets increases with time, as a result of it number of entries in the logs, which can quickly grow to unmanageable size. Automating the analysis of server logs is essential to allow using the logs as a proactive administrator tool. Log analyzer is 3-tier architecture based software that parses through the log files generated by whichever web server follows the standard web server logging standards. Analyze parsed data and categorize them into meaningful reports that can be read by the user for administration or monitoring purpose. A software application designed to parse a log file, which typically contains raw collected data, and convert it to an easy-to-read and understand form.**

General Terms- Data analysis and parsing

Keywords- Server Log analyzer, log file analysis

I. INTRODUCTION

The proposed system is a three layer system which is used for analyze the all servers at same time of an organization. Automating the analysis of server logs is essential to allow using the logs as a proactive administrator tool. Many organizations fail to take full advantage of the available information because of the high initial cost of programming around the various inconsistencies. These log files can be very large and are very detailed on which files were requested from our web server. Every time a page, or image, or movie, or any other kind of file is downloaded off of your web server, the date/time and IP address of the requestor is logged in the web server's log file.

These files used or referred only for troubleshooting purpose when there is problem/issue and those are very difficult to parse due to size, structure and amount of data.But the log is hard to read and understand; also the number of events is huge. Therefore the system administrator only can consult some dispersible materials, and still unable to make the massive materials the statistics and the analysis. Therefore, the website analyzes the software ability to transform these complex materials information simultaneously also provided the extremely user friendly interface, then the system administrator will be very easy to obtain the statistics or the numeral. System also deals with the multiple servers working around the world so that it fetches the data and give analysis to the administrator without need of accessing each server separately. It can gather the log from all related servers and represent it to the administrator.

II. LITERATURE SURVEY

There are all kinds of logs in web servers, recorded all the events of website. Every record in the event is the web page been browsed, the size of web page, what browser has been used, the duration of browsing, and so on. But the log is hard to read and understand; also the number of events is huge. Therefore the system administrator only can consult some dispersible materials, and still unable to make the massive materials the statistics and the analysis. Therefore, the website analyzes the software ability to transform these complex materials information simultaneously also provided the extremely user friendly interface, then the system administrator will be very easy to obtain the statistics or the numeral. For example the majority of analyses result can demonstrate the graphics, some even can provide the PDF files output and so on. But todays available system are not flexible as some are scripting based or require programing knowledge.

A. About log files

Server like Apache, IIS etc. provides the following important logs : (1) Error logs: records all the error information, including CGI script error messages. (2) Access log: records all the requests for he server, and by the analysis of the log we will obtain many precious information. For example: how many customers have recently browse homepages of the website in past week. (3) Script Log: Records CGI script of input and output data. (4) Rewrite LogRecords: the detailed analysis to explain how the Rewrite engine does transform the request.

Proposed analyzer is given the specific path of these log files so that it will fetch these logs and after parsing store it in the database. Every server has provisions to set up the fields which needs to be record in the log files.

III. SYSTEM FEATURES

A. Specification Acceptance

It may happen that any organization has more than one server for its working. So it is important to analyze the logs of all the servers contained in it. In our system we are going to take the logs from all the server machines of a organization. So that it will show the graphical or tabular data of organization as single server.

In our project we are firstly accepting the server names long with the required specification that the administrator wants on the home page. We are going to analyze the log files of the server specified by the application user that is admin of

server. The analysis result we are going to show as per the specification given by the user on home page. The application user also going to choose whether the representation ofanalysis will be in graphical format or tabular format or any other format and also if he requires the actions over the error logs occurred.

B. Parsing and searching

The logs files are simple text files so they are required to parse and generate tokens from it. These tokens are then inserted into the database. According to the users requirements database is searched and required data is given.

C. Analysis Report generation

After the successful storage to database the main thing is to represent the required data in proper format. There are many libraries in java which can represent the tabular values to a specific graphical format like pie chart or bar chart. So its main feature of our system to represent the analysis to the user in easily readable form.

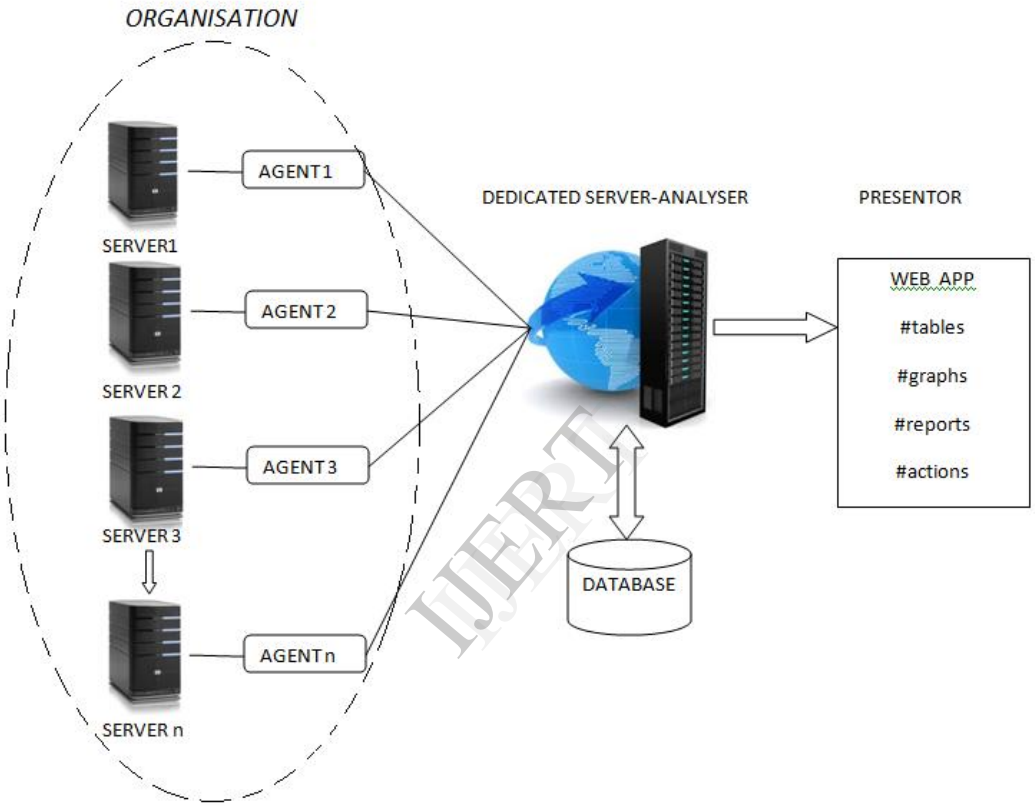

Fig 1: 3-tier system Architecture

IV. ARCHITECTURE

A. System Architectural design

In the above architectural diagram fig. 1, system is consist of 3 tier architecture which is mainly design to handle the group of servers of an organization. It has following 3 main components along with the database:

B. Agent for server organization

In the organization there may be more than one server. So we need to fetch the log files from each server. Agent is a service program of O.S. which fetches the log from each of server and send it to the middleware. It periodically reads the log files and send it to the middleware.

C. Middleware

This is a dedicated server pc which is used to handle and control all backend processes. It contains the database which stores the parsed logs in a table. It takes the input from the agent programs and parse them and store them. Also it takes the request from the front end to search and calculate the data for representation. It will be a system service which continuously runs in background.

D. Presenter

It represent the data in required graphical or tabular form. In this system presenter be a java server pages which controls the users activity. It accept the parameters to analyze the data and send request to middleware. Then middleware

search and calculate values and send it to presenter. By using the ready java classes the data can be represented in a proper graphical form.

E. Communication

In this three tier architecture important thing is connectivity and communication between three components. Remoting service can be used for this communication

F. Structural flow of system

The overall flow of the system goes throughthese stages as shown in fig 2.

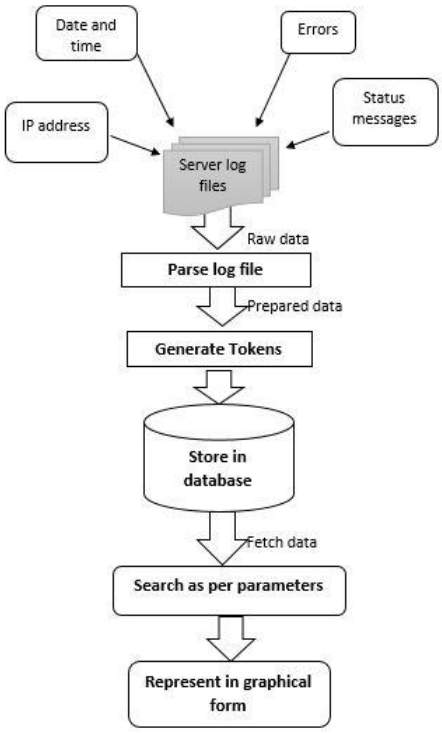

Fig. 2: structural flow of diagram

V. CONCLUSION

Web logs provide vast amounts of information about the use of their websites. Web logs can provide information about which user groups access the website, which resources are viewed the most, and the links users follow to arrive at the site. In order to process the large amounts of data generated by most websites, log analysis software must be utilized. This software produces reports with summary statistics about website. Our program analyzes the information in server's log files and creates detailed reports. It can tell when server was down and for how long, how much bandwidth site has been using. Also this system has main advantages of accessing logs of all servers of an organization at one time so admin has a great advantage of it.

REFERENCES

[1] Dilip sisodia and shrish warma ”Web Usage Pattern Analysis Through Web Logs: A Review”2012 ninth international conference on computer science and engineering.

[2] Nathaphon Kiatwonghong and Songrit Maneewongvatana “Intelli-LogLog : A Real-time Log Analyzer” 201O 2nd International Conference on Education Technology and computer(ICETC)

[3] Wichian Premchaiswadi“ Extracting WebLog of Siam University for Learning User Behavior on mapreduce” 4th International Conference on Intelligent and Advanced Systems (ICIAS2012)

[4] L. Liberti “Log Analysis Software Architecture”

[5] Chen Hu, Xuli Zong, Chung-wei Lee and Jyh-haw Yeh, “World WideWeb Usage Mining Systems and Technologies”, Systemic, Cyberneticsand Informatics Vol. 1 – Number 4.

[6] The Apache Software Foundation, “Log files,” http:// httpd apache.org/ docs/ 1.3/ logs.html, 2010.

[7] Yuan, F., L.-J. Wang, et al. (2003). Study on DataPreprocessing Algorithm in Web Log Mining. Proceedingsof the Second International Conference on Machine Learningand Cybernetics, Wan, 2-5 November 2003.

[8] Thomas Reidemeister, Miao Jiang and Paul A.S. Ward (2011). Mining Unstructured Log Files for Recurrent Fault Diagnosis 12th IFIP/IEEE 1M 2011: Mini Conference