IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Log analytics using ELK stack on Cloud platform

Sunny Advani[1], Meghna Mridul[2], Prof. S. R. Vij[3], Manil Agarwal[4], Loya Palak A.[5], Kasturkar Sanketa S[6]

Softantra PVT. LTD., Pune, India[1]

Computer Department, MAEER'S MIT, Kothrud, Pune, India[2, 3, 4, 5, 6]

Abstract: Log analytics using ELK stack is an implementation of Logstash for extracting and indexing, Elasticsearch for searching and kibana for visualizing which are combined together to perform operation on logs. Elasticsearch is a Lucene based search server. Elasticsearch provides a distributed, multitenant-capable full-text search engine. Logstash tool to collect, index, and forward events and log messages. Kibana is used to represent the data in a graphical and statistical way. In this system we use the logs from Elastic Load Balancer, a service of Amazon Web Servers with the help of Logstash. Then these logs are indexed via logstash and represented through Kibana.These logs are properly formatted and ready for information retrieval. Thus,a system admin can check the behaviour of the whole server or a particular component by going through the logs and analyzing the system status. This will help in identifying the fallback of the server and the system administrator can take necessary actions to resolve the problems if any.

Keywords: Lucene, Amazon Web Servers, Elastic search, logstash, kibana.

1. INTRODUCTION

Today's companies produce large chunks of data and there is no easy way to analyze this data.This data which is generated every single minute is useful to the organizations to analyze their business and develop business strategies. There should be some mechanism through which it becomes easy to analyze such huge amount of data with less efforts.This project serves the purpose of designing a system to analyze this data and provide interface using which the data can be interpreted in an efficient way. The objective of this project is do real time indexing and visualization of logs generated by various servers like web server, application servers. For developing such a system which does log analytics from a huge volume of data received from different servers and applications, three tools which are the important part of this system are used. The three tools are logstash, elasticsearch and kibana. These tools do different activity. Different plugins are used to push the data and this data is collected using logstash along with indexing the data which is then passed to the elasticsearch for searching and querying, kibana is used to display the result of the query processed by the elastic search in a graphical way.

2. BLOCK DIAGRAM

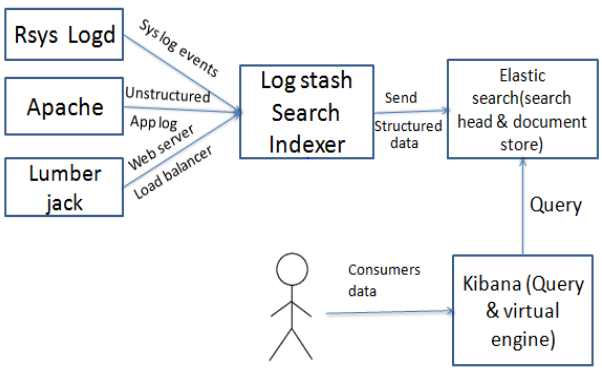

Figure 2.1: Block diagram of the proposed system

Fig 2.1 shows block diagram of proposed system. It consist of Logstash search indexer, elastic search and, kibana. The details of each part will be described in following section.

3. SEMANTIC DESCRIPTION

3.1 ELASTICSEARCH

Elasticsearch is a search engine which is used to store the data given to it by the logstash after indexing. It works as a search engine with real-time analytics. It searches, sorts, filter and analyzes the data. It can work on structured and unstructured data.It not only stores them but also queries the contents of each document as per the user requirements. The user passes its query to the elasticsearch and then processed result is displayed using Kibana.

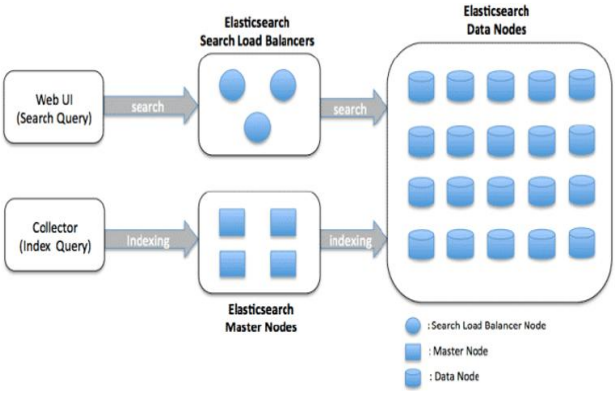

Figure 3.1.1: Components of Elasticsearch

3.2 LOGSTASH

Logstash is an open source data collection engine. It also does indexing on large amount of data or logs that are collected from different servers. This data is passed to the elasticsearch for further query processing. It can dynamically unify data from variety of sources and normalize the data into framework of user's choice. It is a

DOI 10.17148/IJARCCE.2016.5413

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

tool to collect, process, and forward events and log messages. Collection is accomplished via configurable input plugins including raw socket/packet communication, file tailing, and several message bus clients. It does the four main tasks such as parsing the data and logs, extracting the data and logs, managing the logs and structuring it. We get server logs event viewer logs and application custom logs.

- Logstash shipper ships logs to logstash.
- Logstash processes them and does indexing.
- Logstash insert into Elasticsearch.
- Kibana exposes a web interface to Elastic search data.

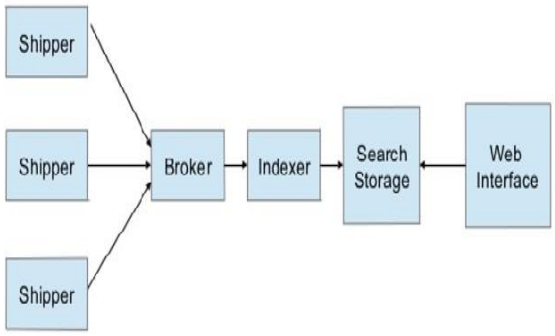

Figure 3.2.1: Logstash architecture

3.3 KIBANA

Kibana is an open source data visualization platform that allows you to interact with your data through stunning, powerful graphics that can be combined into custom dashboards that help you share insites from your data far and wide. This tool does the tasks such as exploring, visualization and discovering data. Depending upon the query and the JSON response the result is generated. Kibana creates tables, graphs, pie charts etc. Thus, kibana does easy representation of large volumes of data and provides analytics. Our system using these tools helps different organizations generating huge amount of data in managing, storing and provides understanding, representation of data in easy graphical forms. It also converts complex and unstructured data to structured data and indexing for its easy representation.

Figure 3.3.1: Kibana overview

4. DESIGN OF PROPOSED SYSTEM

In this section we will describe design of our proposed ELK system.

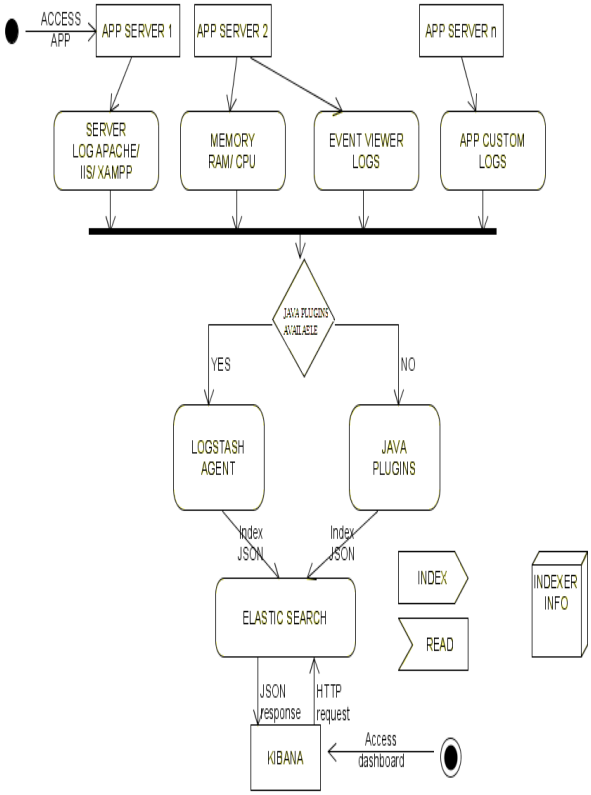

Figure 4.1: Working of ELK stack.

The system consists of three components, as presented in fig. 4.1. They are:

(1)Elacticsearch
(2)Logstash
(3) Kibana

The system consists of three components as mentioned above. Logs are captured from different servers these logs are collected using different logstash agents and given to the logstash for indexing and sorting. Different logstash plugins are used to push the data into the logstash agent. These logstash agents pass the indexed data to the elasticsearch for searching and query processing. User can hand over their queries to elasticsearch using kibana through HTTP request and after processing the query the Json response is created and given to kibana. This result would be displayed on kibana. Result is basically in the form of graphical representation such as pie charts, bar graph, line graphs, histogram due to which reading and understanding of data becomes easier for the end users. Different results can be combined together to form dashboards.

5. RESULT AND DISCUSSION

In this system basically the ELK stack is used which is implemented on cloud platform. Logs from different servers are collected and given to the logstash which does

indexing sorting of these logs which was otherwise difficult. This indexed data when passed to elastic search can then be used for query processing. This query is given by the users to the elasticsearch and results are displayed on Kibana in a easy graphical representation. These results are combined together to form dashboards.

6. CONCLUSION

This system provides log analytics which is based on ELK stack using the cloud platform. As companies produce large chunks of data and there is no easy way to analyze this data. Elasticsearch, logstash and kibana are the main tools of this system which provides an easy way to analyze the data. This system using the above three tools to help different organizations generating huge amount of data in different tasks such as managing, storing, and analyzing and provides understanding, representation of data in easy graphical forms. It also converts complex and unstructured data to structured data and indexing for its easy representation which helps the end user in understanding the data with less effort. These results are finally displayed on dashboards.

7. FUTURE ENHANCEMENT

The project can be extended to some improvements such as:

•Developing a mobile based application for alerts and notification of the performance of the server.
•Requiring minimum physical space or storage area.
•Handling data of big scale companies
•Creating customized tools for extracting logs from application servers and indexing them.

REFERENCES

[1] Jun Bai , ”Feasibility analysis of big log data real time search based on HBase and Elasticsearch”, 2013 Ninth International Conference on National Computation (ICNC),Beijing,China.
[2] Oleksii Kononenko, Olga Baysal, Reid Holmes, and Michael W. Godfrey,” Mining Modern Repositories with Elasticsearch”, School of Computer Science University of Waterloo, Waterloo, ON, Canada okononen, obaysal, rtholmes, migod@uwaterloo.ca.
[3] Bhupendra Moharil,Chaitanya Gokhale,Vijayendra Ghadge, Pranav Tambvekar, Sumitra Pundlik, Gaurav Rai, “Real Time Generalized Log File Management and Analysis using Pattern Matching and Dynamic Clustering”, International Journal of Computer Applications,April 2014.
[4] Ajitpal Singh,Horacio Gonz `I lezˆaV `I lez,” Hierarchical Multi-Log Cloud-Based Search Engine”, 2014 Eighth International Conference on Complex, Intelligent and Software Intensive and Systems, Dublin, Ireland. www.ncirl.ie/cloud.
[5] C. Bhadane, H. A. Mody, D. U. Shah, P. R. Sheth,”Use of Elastic Search for Intelligent Algorithms to Ease the Healthcare Industry”, International Journal of Soft Computing and Engineering, January 2014.

DOI 10.17148/IJARCCE.2016.5413