Log Real-time Management Scheme Based on LEK

Xiafei Lei[1, a], Zhe Wang[2, b], Yuzhen He[3, c]

[1]Department of Computer Science and Technology, Jilin University, Changchun, 130000,China

[2]Department of Computer Science and Technology, Jilin University, Changchun, 130000,China

[3]WHY-E Science and Technology Co.,Ltd, Changchun,130000,China

[a]email: leixiafei1207@163.com, [b]email:wz2000@jlu.edu.cn,[c]email:heyuzhen@why-e.com.cn

Keywords: Logstash, Elasticsearch, Kibana, log management, LEK

Abstract. The growing importance of log analysis and the management play an important role in the era of big data especially. With mass log information increasing, it's difficult to extract the needed information quickly, What's more, most cases are related information. Logstash can solve this problem suitably because of a built-in logging service.It can be combined with help from many log slaves to find data easily . Strictly speaking, Logstash alone are unable to rationalize and centralized the log files. This article integrated Logstash, Elasticsearch and Kibana into an organic whole, which is applied to an insurance business trading system LEK. It implements collection, query, polymerization and visual function of the transaction log, and form the unified convenient log management and analysis tool.

Introduction

Log analysis and monitoring, with rising attention, occupies a very important position in the system development especially, the more complex the system is, the more important log analysis and monitoring becomes. We can query log details according to the keyword, monitor the operation condition of the control system, capture abnormal data, and trigger notifications automatically according to the different level; Also can carry out statistical analysis, performance analysis, etc., such as the number of called interface, the execution time, the success rate, etc. As the coverage of big data , many people started data mining based on log, opening up another piece of heaven and earth. But there are some problems in the log analysis, such as mass log data is scattered across multiple systems and querying speed is slow. The most important is that the data is difficult to get in real-time. At present, for some common heavyweight open source Trace system (such as facebook scribe and twitter zipkin, etc.)[1], configuration and deployment are relatively too complicated, we need a lightweight solution. This article combined Elasticsearch, Logstash, and Kibana (named LEK) together to form a log analysis system, realizing the log collection,query and visual display. The structure is as follows: the first part is the introduction, describes the background of ELK; The second part profiles combination member of LEK; The third part is environment deploymen t and system test; The fourth part is the test results and conclusions.

LEK Introduction

Logstash - Log Pipeline

Logstash is the core component of log management system in this paper, it is a completely open source collecting and managing log tools. Besides view the log, its compon ent architecture support to manage log flow of the different servers through a proxy, and sent to storage eventually. Logstash is not geared to the needs of the client and the server of a single Java library,but a configuration file (named xx. conf) covering all the services.It mainly include input, a filter and the output,the three functions of LogStash are performed according to the configuration information, each Logstash instance is customized according to the requirements of its role in the overall archite cture[2]. On the function allocation, Logstash is divided into the index and the agent, the agent is responsible for

monitoring and filtering log, index is responsible for collect ing the log and to do search in ElasticSearch.

Elasticearch - Query Engine

Elasticsearch(ES), is a open source searching and analytical engine based on Lucene, and supports cloud service[3]. Through it we can dig deeper into the data easily by amplifying and narrowing the range of search and analysis, most important is real-time[4]. In this article ES is used as "log database" and a view of back-end system. Unique characteristics of ES make query log flexibly, it can realize the filter query, regular query, nested query, etc. Filter is faster than the query for log, besause can automatically be cached in the filter without performing score. ES splits the field in accordance with the separator by default, but some fields can't participles in log (for example url), we can set this type of field "not_analyzed" attribute. In a word,ES has the features of high availability, high scale and real-time processing.

Kibana - Query Visualization

Kibana, a dashboard of Elasticsearch search and the analysis, is based on the browser, using Apache open source protocol[5]. Simply said, it is a web interface of log analysis for Logstash and Elasticsea rch. It can analyze and compare the important data and visualize query results with a rich graphics. In the article, Kibana focus on real-time monitoring and analyzing problem, can combine different time sequence from many dimensions for an event, cooperating with the different conditions of multiple queries.Namely use it as the tripartite interface of Logstash and ElasticSearch, it’s easy to read because of beautiful web shell, also can give a person with aesthetic feeling.

Integration Principle of LEK

Specific workflow of LEK is simple: first, start ElasticSearch and Kibana respectively, and then through the Logstash configuration file, whose function is integrating LEK, to start the Logstash. Specified input, filter, and the output in the configuration file. Logstash agent monitors and filters log, and sent filtered log to Redis (Logstash used Redis as a message broker in the default configuration , which is an open source Key-Value database, it is not for storage, but used to queue in front of the index log); The logstash index collects log and gather to ElasticSearch, then customized search by using ElasticSearch, including complex nested query, regular query or filter query, etc; Finally combining with Kibana to visualize query results, flow chart is illustrated in Fig.1:

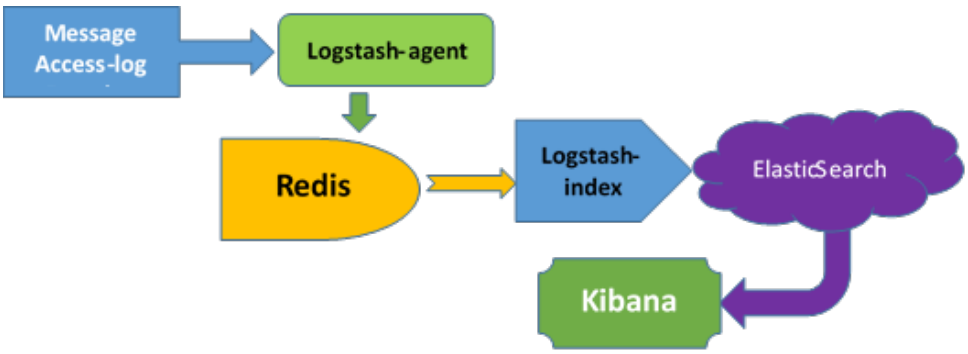

Fig.1. Schematic Diagram of LEK

The experiment part

Experimental Content and Data

This experiment is based on the transaction log of a certain insurance business system, first the system is deployed in Tomcat, and then use Logstash monitor log folder of the native Tomcat, set Tomcat logs is input in the configuration file of Logstash, and output is Elasticsearch. Logstash is equivalent to a log line at this time, the log is sent to Elasticsearch and stored, and then through Kibana customize query and visualize query results. The experiment use the July transaction log has many fields, mainly discuss type, channel ,time ,result and Gateway of transaction ,and the relations among them, 4009589 in total.

Experiment Process:

1) Install the JDK 1.7.0, Elasticsearch - 1.4.4, Logstash - 1.4.2, Kibana - 4.0.1.

2)Deploy the trading system in Tomcat and Start the Tomcat .

3)Start the Elasticsearch and Kibana correctly.

4)Start the Logstash:Through CMD into logstach bin directory, enter " logstash agent - f test.conf " (test.conf set the input for the Tomcat path,and the output for Elasticse arch).

5)Input " I'm testing Logstash" in Logstash,and refresh Kibana to check if there is an index named logstash*,it's successfully integrated if you find the "I'm testing Logstash" in index logstash*.

6)Monitor Elsticsearch ,check whether Logstash send the the transaction log to it or not.

7)From the 4009589 records,begin to find the complex relations by Kibana:

a) The five transaction types of top 10 channels in July,and result in Fig.2.

b) Appoint the gateway(GW006),statistic top 5 of processing request channels and top 5 types of transactions in corresponding channel in July,and result in Fig.3.

c) Compare each channel dealting with the number of deals every day in July,and result in Fig.4.

Experiment Result:

We do three experiments to test the effectiveness of LEK,from simple aggregation query,situation comparison ,to complex relation mining.As the Fig.2, Fig.3 and Fig.4.

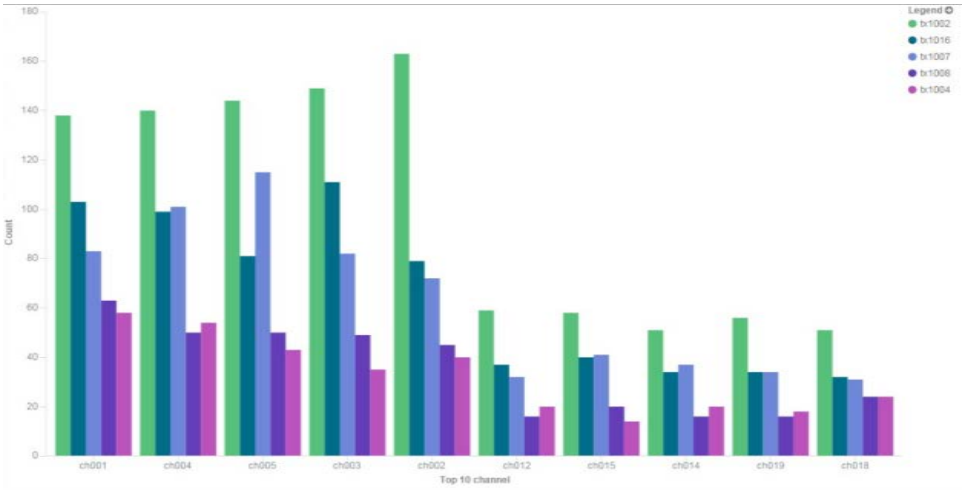

Fig.2.Result of a

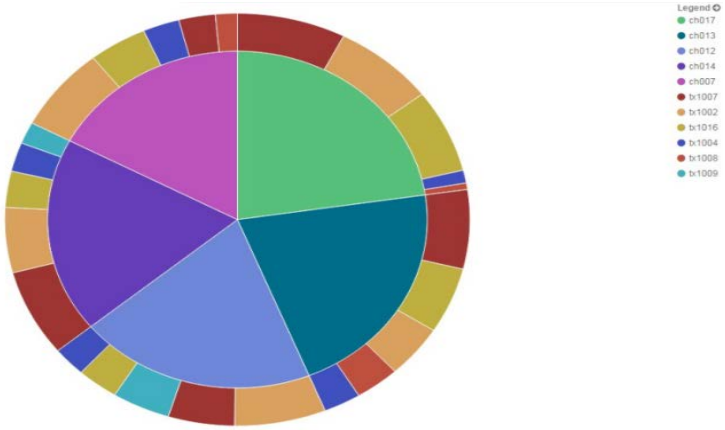

Fig.3.Result of b

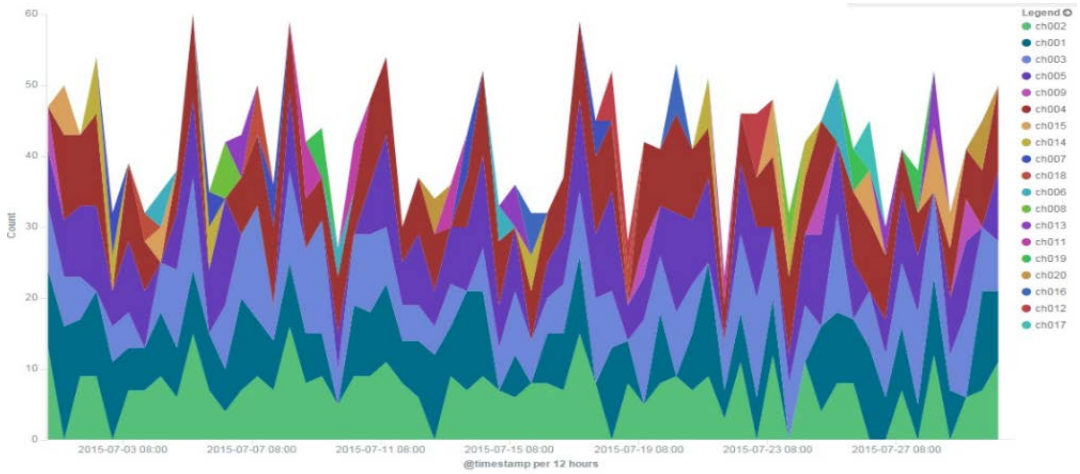

Fig.4.Result of c

Summary

Monitoring and troublehooting is the most common use case of log management. Log is also a very good source of analysis, as the experiment above we can use log to analysis the potential relationship among different fields.By LEK,we know the complex situation of transaction and sub-aggregation based on certain condition, such as channe17 is the most request at gateway(GW006). In addition,When a transaction has problems, the log data is often the best source of information.If you follow the “checklist” your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

Acknowledgements

This work was financially supported by the National Science and Technology Support Projects-Jilin Financial Institutions Application Demonstration Project(Project No.20 13BAH07F05).

References

[1] JunBai, HeBinGuo. Software Integration Research of Large-scale Logs Real-time Search Based on ElasticSearch[J].Journal of Jilin Normal University (Natural Science Edition),2014(2).

[2] James Turnbull,The Logstash Book[OL].http://logstash.net.2014.01.

[3] S Bagnasco,D Berzano Monitoring of IaaS and scientific applications on the Cloud using the Elasticsearch ecosystem.[J].Journal of Physics: Conference Series,2015(6).

[4] Junjie Chen, Guofan Huang. Reconstruct library search engine based on Elasticsearch[J]. Information Research,2015(11).

[5] DouShi San, Kibana user guide in Chinese[OL].http://kibana.net.2014.06.