Analytical monitoring platform based on ELK for Beidou system log

Meng Fangyuan[1], Wei Tao[1,2]

1. Beijing Satellite Navigation Center, Beijing, China

2. Sichuan University, Chengdu, China

1. 185427910@qq.com, 2. 675279854@qq.com

Abstract: Nowadays, satellite navigation has been applied to all aspects of people's livelihood, infrastructure and national defense because of its high precision, all-weather and non-ground station restrictions. How to ensure the stability of the system related to the national economy and the people's livelihood. In the large-scale satellite navigation and positioning system, the analysis and processing of the massive log and the monitoring of the system status are the key factors to ensure the stability of the system. The existing system through the log troubleshooting issues are through the remote login server to the appropriate directory query,time-consuming and laborious. This paper presents a log analysis and monitoring system based on Elasticsearch, Logstash and Kibana, which is used to collect and retrieve massive log and system information in real time. This system can find, locate and deal with system abnormity more quickly and improve the usability of satellite system. This paper describes the framework of the system, the function and working principle of each component, and describes in detail the process of collecting, filtering and indexing the log and system information to the final query and display.The results show that the method can be used to improve the system in the event of failure through the log for investigation and processing speed.

Keywords: System monitoring; log analytical; ELK

基于ELK的北斗系统日志分析监控平台

孟方园[1]，魏 涛[1,2]

1．北京卫星导航中心，北京，中国，100094

2．四川大学，成都，中国，610065

1. 185427910@qq.com, 2. 675279854@qq.com

【摘要】现如今卫星导航由于其高精度、全天候及不受地面基站限制的特点已经被应用到了民生、基建及国防的方方面面。如何保证系统的稳定性关系到国计民生。而在大规模的卫星导航定位系统中，对海量日志的分析和处理及系统状态的监控是保证系统稳定的关键因素。现有系统在通过日志排查问题时均是通过远程登录服务器到相应目录查询，费时费力。本文提出了一种基于Elasticsearch，Logstash和Kibana的日志分析监控系统，用于对海量日志和系统信息进行实时采集和检索，使发现、定位和处理系统异常更为快速，从而提高卫星系统可用性。文中主要阐述了该系统的框架、各组件的功能及工作原理，详细描述了从日志及系统信息的收集、过滤及索引到最后的查询展示的过程。结果表明使用该发方法可以有效的提高在系统发生故障时通过日志进行排查和处理的速度。

【关键词】系统监控；日志分析；ELK

1 引言

大数据时代，资源的种类和数量也越来越多，这标志着系统建设的日趋完善，同时也意味着运维管理将面临更大的挑战。北斗卫星导航系统[1]在不断迈向全球化的进程中，系统规模不断扩大，业务功能日益完善，与此同时各种日志信息同时也呈数量级的增长。日志中包含了系统运行过程中的许多有用信息，包括中间结果信息、告警信息等。通过对日志的分析可以迅速发现系统瓶颈，优化系统性能及修复系统漏洞。日志主要包括系统日志、应用程序日志和安全日志。系统运维和开发人员可以通过日志了解服务器软硬件信息、检查配置过程中的错误及错误发生的原因。经常分析日志可以了解服务器的负荷，性能安全性，从而及时采取措施纠正错误。

2 ELK简介

通常，日志被分散的储存不同的设备上。如果管理数十上百台服务器，需依次登录每台机器的查阅日

志，这样的方式较为繁琐和低率。当务之急我们使用集中化的日志管理及统计和检索。一般情况下使用grep 、 awk 和 wc 等 Linux 命令能实现检索和统计，但是对于要求更高的查询、排序和统计等要求和庞大的机器数量依然使用这样的方法难免有点力不从心。

近年来利用ELK技术搭建日志分析平台的方式有不少研究，如陈建娟，刘行行研究的基于Kubernete的分布式ELK日志分析系统[2]，吕佳讨论的基于ElasticSearch的分布式日志搜索系统设计[3]，李祥池对基于ELK和Spark Streaming的日志分析系统设计与实现的研究[3]，均讨论了ELK在日志分析领域的特点及可行性。

本文采用ELK实现日志收集分析，ELK能在大量日志中快速准确定位故障，适合应用级别的实时监控和重要核心服务的报警。ELK 是一套流行的一体化日志处理平台解决方案，由分布式搜索引擎ElasticSearch、日志采集解析工具 Logstash、分析可视化平台Kibana组成，提供日志收集、处理、存储、搜索、展示等全方位功能。

Elasticsearch 是一个强大的具有搜索功能的非结构化数据库，作为一个开源分布式搜索引擎，它的特点有：分布式，零配置，自动发现，索引自动分片，索引副本机制，restful风格接口，多数据源，自动搜索负载等。

Logstash 是 Ruby 编写的一款分布式日志收集系统，作为一款轻量级的日志搜集处理框架，可以方便的对分散的、多样化的日志进行搜集，并进行自定义的处理，然后传输到指定的位置，比如某个服务器或者文件。Logstash使用管道方式进行日志的搜集处理和输出。在logstash中，包括了三个阶段:输入input ->处理filter -> 输出output。

Kibana 提供日志分析的 web可视化界面，利用 ElasticSearch 搜索功能，以秒为单位可视化数据，支持 Lucene 的查询字符串的语法和Elasticsearch的过滤功能。

3 平台架构

在需要收集日志的服务器上安装Beats组件，由Beats收集日志及系统信息，并统一发送给主节点上的Logstash，Logstash分析、过滤日志数据后发送至Elasticsearch存储，并由Kibana最终将数据呈现给用户，架构如图1所示。

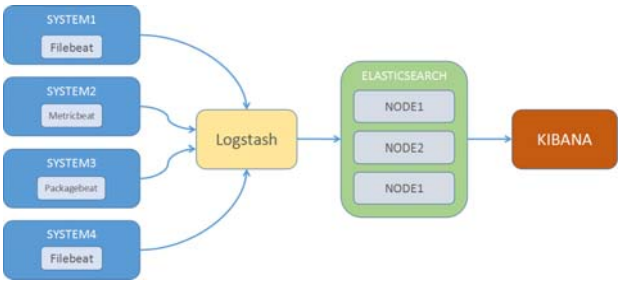

Figure 1. System struts

图 1.系统架构图

Beats platform目前包含有Packagebeat、Metricbeat和Filebeat等产品，均为Apache 2.0 License，同时用户可根据需要进行二次开发。该方案中Beats所消耗的系统资源较少，并且针对Logstash和Elasticsearch可进行集群配置，以分担负荷。日志经过Logstash分析、过滤后发送给远端服务器上的Elasticsearch进行存储，Elasticsearch将数据以分片的形式存储并提供多种API供用户查询，操作。用户亦可以更直观的通过配置Kibana Web Portal方便的对日志查询，并根据数据生成报表。

4 系统组成

4.1 日志获取组件

日志的获取是通过安装在服务器上的各类Beats组件实现的，包括Filebeat、Packectbeat和Metricbeat等。本平台主要使用Filebeat组件。

Filebeat是一个开源的文件收集器，主要用于获取日志文件。Filebeat安装在服务器上作为代理来监视日志目录或指定的日志文件，并将日志转发至Logstash进行解析或ElasticSearch进行索引，如图2所示。

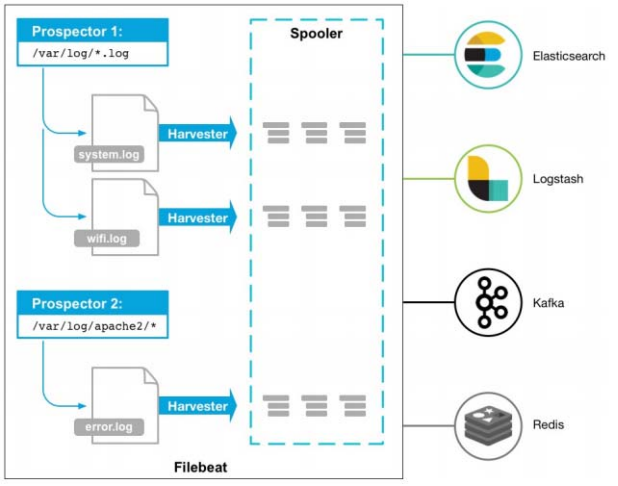

Figure 2. Filebeat flow chart

图 2.Filebeat 流程图

4.2 日志处理组件

Logstash作为日志处理的关键一步，通过接收各服务器上Beats传送过来的日志数据，经filter后最终将格式化的日志输出至elasticsearch进行索引，并每天会按照默认格式logstash-YYYY.MM.DD来建立索引。在午夜(GMT)，Logstash自动按照时间戳更新索引。我们可以根据追溯多长时间的数据作为依据来制定保持多少数据，当然你也可以把比较老的数据迁移到其他的地方(重新索引)来方便查询。

Inputs,Outputs,Codecs,Filters构成了Logstash的核心配置项。Logstash通过建立一条事件处理的管道，从你的日志提取出数据保存到Elasticsearch中，为高效的查询数据提供基础。

input 及输入是指日志数据传输到Logstash中。其中常见的配置如下：

- file：从文件系统中读取一个文件，很像UNIX命令 "tail -0a"
- syslog：监听514端口，按照RFC3164标准解析日志数据
- beats：监听5044端口，接收远端服务器传来的日志数据

Fillters 在Logstash处理链中担任中间处理组件。他们经常被组合起来实现一些特定的行为来，处理匹配特定规则的事件流。常见的filters如下：

- grok：解析无规则的文字并转化为有结构的格式。Grok 是目前最好的方式来将无结构的数据转换为有结构可查询的数据。有120多种匹配规则，会有一种满足你的需要。
- mutate：mutate filter 允许改变输入的文档，你可以从命名，删除，移动或者修改字段在处理事件的过程中。
- geoip：添加地理信息(为前台kibana图形化展示使用)

outputs是logstash处理管道的最末端组件。一个event可以在处理过程中经过多重输出，但是一旦所有的outputs都执行结束，这个event也就完成生命周期。一些常用的outputs包括：

- elasticsearch：高效的索引及保存日志数据，并且能够方便和简单的进行查询。
- file：将日志数据保存到文件中。

codecs 是基于数据流的过滤器，它可以作为input，output的一部分配置。Codecs可以帮助你轻松的分割发送过来已经被序列化的数据。流行的codecs包括json, msgpack, plain(text)。

- json：使用json格式对数据进行编码/解码
- multiline：将汇多个事件中数据汇总为一个单一的行，如：java异常信息和堆栈信息

一个完整的logstash配置如图3所示：

```
input {
     beats {
        port => 5044
     }
}

filter {
  grok {
    match => { "message" => "%{COMBINEDAPACHELOG}" }
  }
  date {
    match => [ "timestamp" , "dd/MMM/yyyy:HH:mm:ss Z" ]
  }
}

output {
  elasticsearch { host => ip:port }
  stdout { codec => rubydebug }
}
```

Figure 3. Logstash config

图 3. Logstash 配置

在终端运行命令：“nohup bin/logstash -f logstash-filter.conf &”，即可使logstash按logstash-filter.conf文件中的配置在后台运行。其中grok插件提供了多数常见的正则匹配，同时根据日志的格式还可定义自己需要的正则匹配。经过filter处理后的日志文件被格式化为JSON格式，用以输出到ElasticSearch进行索引。

4.3 日志索引组件

日志的索引工作是由Elasticsearch完成的，通过在logstash的output中指定ElasticSearch集群的任意一节点IP和端口即可将日志导入并完成索引。

以单文件的静态层面看，每个全文索引都是一个词元的倒排索引。以在线动态服务的层面看，要做到实时更新条件下数据的可用和可靠，就需要在倒排索引的基础上，再做进一步的处理。Lucene的做法是将新收到的数据写到新的索引文件里，每次生成的倒排索引叫做一个段（segment）。

4.4数据可视化组件

该平台日志数据的可视化使用开源项目Kibana。仅需下载对应版本的二进制包，解压文件，并执bin/kibana(Linux/MacOSX)或bin\kibana.bat（Windows）即可。在Setting界面指定一个或多个index pattern，一般由logstash导入的数据索引可用logstash-*进行匹配。选择一个包含了时间戳的索引字段，可以用来做基于时间的处理。Kibana会读取索引的映射，然后列出所有包含了时间戳的字段。

Kibana可以根据导入的日志数据生成表格、柱状图、饼状图等，并且利用Kibana的仪表盘自由的组合这些图表，方便监测各类日志数据，如图4，图5所示。

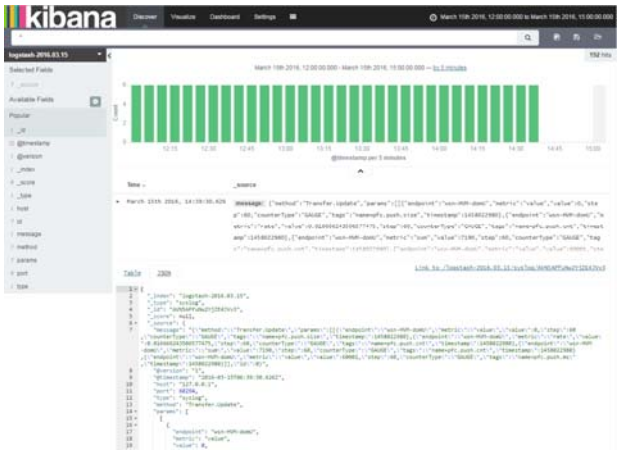

Figure 4. Kibana visualization

图 4. Kibana 视图

Figure 5. Kibana Dashboard

图 5. Kibana 仪表盘

5 结论

本文介绍了基于ELK的日志处理平台在北斗日志分析处理中的应用。为系统运维提供了有效的应对方法。通过应用该平台，提高了排查和处理系统故障的效率。

References (参考文献)

[1] Tan Shusen, The Development and Consideration of Compass Satellite Navigation System, Journal of Astronautics, 2008.02. 谭述森，北斗卫星导航系统的发展与思考，宇航学报，2008.02.

[2] Cheng Jianjuan, Liu Xingxing, A Distributed ELK Log Analysis System Based on Kubernete. Electronic Technology & Software Engineering, 2016.2:211-214. 陈建娟，刘行行，基于Kubernete的分布式ELK日志分析系统，电子技术和软件工程，2016.2:211-214.

[3] Lv Jia, A distributed log-search system based on Elastic Search, 2013.09.吕佳，基于ElasticSearch的分布式日志搜索系统设计，2013.09.

[4] Li Xiangchi, Design and Implementation of Log Analysis System Based on ELK and Spark Streaming, Electronic Science and Technology, 2015.2(06):674-678. 李祥池，基于ELK和Spark Streaming的日志分析系统设计与实现 [J]. 电子科学技术，2015.2(06):674-678.