ELK 日志分析平台在电子商务系统监控服务中的应用

Application of ELK Log Analysis Platform in E-commerce System Monitoring Service

联通系统集成有限公司　周　映　韩晓霞

摘　要　针对运维服务人员因海量的应用日志、系统日志、错误日志等分散存储在不同设备里而面临异常监控发现不及时、日志分析效率低下、统计检索繁琐等问题，通过研究 ELK(由弹性搜索、日志事件采集、统计展示三个开源免费工具组成) 的工作原理，完成 ELK 日志分析平台的安装部署，从而提高监控服务质量和效率，实现异常首发。该平台不仅配置简单，而且完全开源免费。该设计也能为运维服务人员直观展现各类日志情况，便于对监控数据分析利用，从而提升运维效率与质量。

关键词　运维　日志分析平台　效率提升　监控服务　ELK

Abstract: For the problems as abnormality unable to be found timely, low efficiency of log analysis, statistical retrieval sophisticated faced by operation and maintenance service personnels because of distributed storage of massive application log, system log and error log in different devices, by studying the the working principle of ELK (formed by three open source tools for free - elastic search, log event collection, statistics show), deployment and installation of ELK log analysis platform are completed to improve the monitoring service quality and efficiency, thus realizing anomaly to be detected firstly and timely. This platform is not only simple in configuration, but also completely free and open source. This design can also directly show all kinds of log to operation and maitenance service personnels, facilitating them in analysis and use of monitoring data, so as to improve operation and maitenance efficiency and quality.

Keywords: operation and maintenance; log analysis platform; efficiency analysis; monitoring service; ELK

1 引言

近年来，互联网在人们的生活中扮演着越来越重要的角色，随着用户访问量的飞速增加，为之提供支撑的服务器端存放的日志信息量也随之剧增。通过传统的单台登陆方式进行监控和运维服务势必无法准确及时地找出海量日志信息里的关键信息。

运维服务人员通常采用集中化日志管理 (如 syslog)，将所有服务器上的日志收集汇总从而实现日志分析和统计。但当服务器数量巨增，且查询、排序及统计要求提高时，该方法便显得力不从心。而 ELK 日志分析平台 (以下简称为 ELK) 使运维人员从浩如烟海的日志信息中轻松准确地监控及维护所需关注的信息及实现日志统计分析等成为了可能。

搭建 ELK 可以有效地解决上述问题，运维服务人员可以根据维护需求定制日志展示的格式和要求，ELK 会根据运维服务人员的要求，将符合特性要求的日志进行汇集，通过前端页面进行图形展示。维护和监控人员可以更简单快速地通过直观的日志统计展示情况了解服务器的资源使用情况、检查配置过程中的错误及错误产生的原因。

2 ELK 工作原理

ELK 由 Elasticsearch(弹性搜索)、Logstash(日志

事件采集)、Kibana(统计展示) 三个开源免费工具组成。

图 1 显示了 ELK 的工作原理。

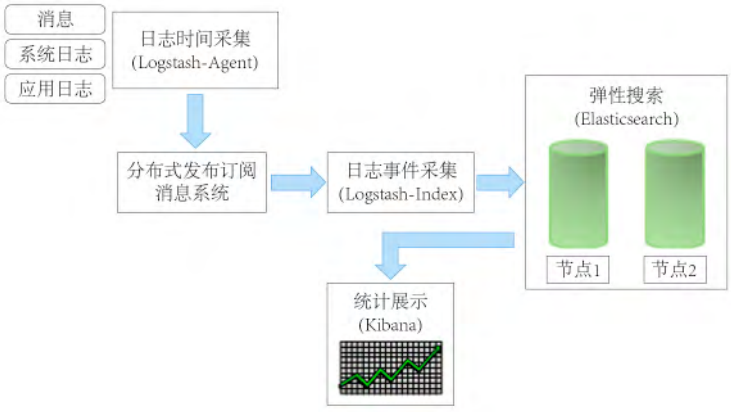

图 1　ELK 工作原理图

Elasticsearch[1] 是基于全文检索引擎工具包 (Lucene) 构建的开源且满足 Restful 设计风格的搜索引擎。其主要是提供一个分布式框架扩展 Lucene，从而实现大数据量、分布式的搜索功能，设计用于云计算中，能够达到实时搜索、稳定、可靠、快速、安装使用方便。其实现思想简单，将大数据量分而治之，对每一份进行索引、检索，最后将每份结果合并返回。

Logstash[2] 是一个完全开源的工具，可以对日志进行收集、分析，并将其存储供之后使用 (如搜索)，在一个典型的 ELK 场景下，Logstash 在其过程中担任搬运工的角色，它为数据存储、报表查询、日志解析等行为创建了一个功能强大的管道链，并提供多种多样的组件，让使用者可以轻松实现强大的功能。

Kibana[3] 也是一个开源和免费的工具，可以为 Elasticsearch 和 Logstash 提供友好的日志分析网页展示界面，并可以帮助汇总、分析和搜索重要数据日志。

如果在需要收集日志的所有服务器上部署 Logstash，它的监控过滤角色会不断地读取需要的日志信息，每当读到新的日志信息后，就将信息传送到分布式发布订阅信息系统的队列上，对于队列上未做处理的日志，Logstash 的日志收集角色会进行接收和分析，分析后存储到 Elasticsearch 中进行搜索，再由统一的 Kibana 进行网页日志界面的展示。

3 ELK 在监控服务中的应用

本次实施中，以某电子商务系统为例，通过对 ELK 的安装部署，提升日志分析效率及准确性。

3.1 软硬件环境

本次实施需配置 38 台刀片机：配置 20 台刀片机用于部署应用服务，7 台刀片机用于部署分布式发布订阅消息工具，10 台刀片机用于部署弹性搜索工具，1 台主机用于部署统计展示工具，每台机器均安装 64 位 Linux 操作系统，内存为 32 GB，自带 400 GB 硬盘。

下载 Java 虚拟机和 Logstash、Kibana、Elasticsearch 工具的最新版本，执行安装后对这些工具配置文件进行路径、应用服务器地址及端口等配置以便参与日志的收集过滤、分析和展示。

3.2 文件配置

(1) 进入 Logstash 目录，创建配置文件 (Logstash.conf)，并修改配置文件里的输入 (input) 和输出 (output)，将日志路径添加至 input(可支持多路径配置)，将安装有分布式发布消息订阅系统的服务器及端口和 ElastichSearch 访问地址及端口添加至 output(可支持多服务器)。

(2) 创建启动 Logstash 的启动脚本，并配置监听文件，将 Logstash 的路径和启动脚本的路径添加至该监听文件中。

(3) 创建 Kibana 的监听配置文件，设定访问端口并将 Elasticsearch 的访问地址及端口添加至配置文件中。

3.3 应用服务

ELK 部署完成后，日志收集后通过特定条件统计展示的效果如图 2 所示。

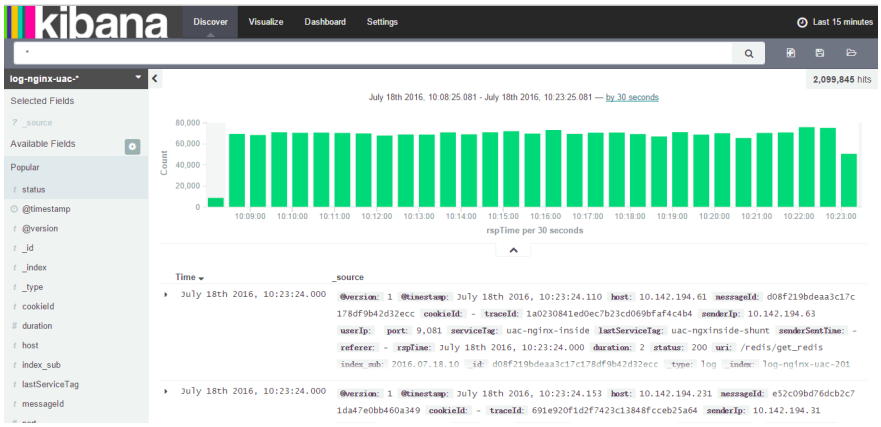

图 2　统计展示图样例

该 ELK 为监控服务提供了单日志查询和统计分析两大功能。

当需要查询某一条日志记录信息时，点击“ Discover(展示) ”按钮，在左侧标签栏下拉列表中选择对应的系统名后展示出所有的标签项，再搜索框中输入具体的标签项和需要查询的值便可完成单日志查询，如查询 A 号码的交易日志，则输入“ 标签项：A ’”完成搜索，如需再查询条件上继续过滤，只需使用“ AND ”或者“ OR ”(需大写) 将筛选条件组合起来。

在日常监控服务中运用最广泛的是统计分析功能，通过对数据的整理筛选及时发现系统异常并解决。Kibana 提供了八种基本图形供选择，当需要绘制系统核心业务走势图时，点击“ Visualize(设计) ”按钮，选择“ vertical bar chart(直方图) ”，首先需要给 X 轴添加时间点击左侧的“ X-Axis ”，在“ Aggregation(集成) ”处选择 ”Date Histogram(日期柱状图) ”，在“ Field(域) ”处选择时间标签。

接下来点击“ Add sub buckets(添加子集) ”，选择“ Split Bars(拆分条) ”，在“ Sub Aggregation(子聚集) ”处选择“ Filters(过滤) ”，因绘制走势需要成功量和失败量，所以需要两个 Filters，点击下方“ Add Fileter(添加过滤) ”添加。

最后添加接口部分，点击下方的“ Add sub-buckets(添加子聚集) ”，选择“ Split Chart(拆分条件) ”，在“ Sub Aggregation(子集成) ”处选择‘ Terms(条件) ”，在“ Field(域) ”处选择接口标签，点击执行便可得出业务走势图。

除走势图外，其他图形结构也使得系统的日常监控服务更加高效、快捷和方便。

4 实现效果

ELK 完成部署后，对该电子商务系统的日志统计分析效率及单数据日志查询效率进行了统计。选取部署前和部署后一天中日志量最大的时段进行数据的操作，对同一条数据日志记录进行查询，并计算出该数据的查询速率。

经分析得出，部署前需要逐个对 20 台应用服务器进行查询，耗时最长可达到 10 min。而部署 ELK 后，通过搜索框中输入相关标签及内容，耗时不到 1 s 便可得出查询结果，且该速率并不会因日志量或者服务器数量的增加而改变。

同时，ELK 收集的应用服务器数据，通过 Kibana 的“ 设计 ”功能，按需求绘制出“ 业务量走势图 ”、“ 错误返回码占比图 ”、“ 网络平均耗时 ”、“ 服务器队列 ”等多元化的监控图，维护人员快速准确地发现了多起该电商系统在业务高峰期的异常日志进程、网络流量异常及日常运行中存在的隐患等。维护人员可以直观地通过图形实时了解系统的性能及资源占用情况，及时提出系统优化建议。尤其在出

现故障时能快速的发现和定位，提升了故障的响应速度和处理效率，缩短了故障历时，将故障影响范围降到最低。大大提升了监控和运维服务的质量。

同时，当统计分析需求变得更复杂繁琐时，如“统计某个时间段，系统错误返回码占比 TOP5”，部署前是无法通过日志查询来实现的，部署 ELK 后，可通过“设计”功能，选择相关标签及展示图表格式，在 1 min 之内得到指定条件的结果展示。

ELK 不但大大提高了日志搜索速率，其在日志统计分析时对监控和运维服务带来的便捷更是传统监控和维护方式无法比拟的，且由于其日志均实时同步至 Elasticsearch 中，在进行日志查询分析时并不会影响到正常的生产服务业务，在提升服务质量的同时还能保证系统的稳定运行。

5 结语

本文针对运维人员和监控人员对日志信息快速查询和统计的需求，基于 ELK 开源免费工具，充分利用现有硬件资源构建了高效的日志分析平台，解决了以往日志分析效率低，耗时长，异常问题定位不准确、发现不及时且过程繁琐等问题。

由此可见，ELK 适用于解决运维服务和监控服务中针对大量服务器的日志快速提取和分析等问题。且该平台除可用于简单的业务数据查询操作外，对超大数据量，较为复杂及需个性化定制的日志分析操作需求等均可引用实现。尤其针对监控服务在整个 IT 服务中担任的故障首发角色起到了关键性的作用，其在监控服务中能快速发现异常日志，为运维服务中的故障定位提供准确的依据，并能通过异常日志的统计分析结论为系统维护优化服务提供依据。

目前 ELK 日志分析平台已经成功运用于通信行业部分电子商务系统中，取得了不错的效果。

参考文献

[1] 拉法乌 . 库奇 , 马雷克 . 罗戈津斯基 , 时金桥 . Elasticsearch: 可扩展的开源弹性搜索解决方案 [M]. 北京 : 电子工业出版社 , 2015: 40-135.
[2] 高凯 . 实战 Elasticsearch、Logstash、Kibana: 分布式大数据搜索与日志挖掘及可视化解决方案 [M]. 北京 : 清华大学出版社 ,2015:50-235.
[3] 饶琛琳 . ELK stack 权威指南 [M]. 北京 : 机械工业出版社 , 2015: 70-335.

(收稿日期：2016-07-01)

(上接第 58 页) 过剖析智能制造 / 工业 4.0 的核心要素和特征，逐步分解为维度、类和域，并通过对域的要求分级，来呈现出为实现各个要素应达到的要求。

三是在具体的维度、类和域的内容上是不同的。能力成熟度侧重制造与智能两个维度，而 4.0 就绪度更侧重于智能技术本身；此外能力成熟度兼顾了流程与离散两个行业的特征，而后者更侧重于离散行业。能力成熟度更适合当前中国制造业的发展国情。

四是在应用推广层面，4.0 就绪度目前已在德国 234 家机械和装备工程企业中进行了初步评价，而下一步我们也将致力于能力成熟度在中国制造业中的应用。

参考文献

[1] Lindner,T&W,Manfred. INDUSTRIE 4.0 READINESS[R]. Frankfurt: VDMA's IMPULS-Stiftung, 2015: 21-25.
[2] 工业和信息化部 , 国家标准化管理委员会 .《国家智能制造标准体系建设指南》[EB/OL].[2016-05-25]. http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057585/n3057589/c4570069/content.html.
[3] 于秀明 , 郭楠 , 王程安 , 等 . 智能制造能力成熟度模型研究 [J]. 信息技术与标准化 , 2016(05): 39-42.

(收稿日期：2016-06-17)