ELK 日志分析

1. 为什么用到 ELK

一般我们需要进行日志分析场景：直接在日志文件中 grep、awk 就可以获得自己想要的信息。但在规模较大的场景中，此方法效率低下，面临问题包括日志量太大如何归档、文本搜索太慢怎么办、如何多维度查询。需要集中化的日志管理，所有服务器上的日志收集汇总。常见解决思路是建立集中式日志收集系统，将所有节点上的日志统一收集，管理，访问。

一般大型系统是一个分布式部署的架构，不同的服务模块部署在不同的服务器上，问题出现时，大部分情况需要根据问题暴露的关键信息，定位到具体的服务器和服务模块，构建一套集中式日志系统，可以提高定位问题的效率。

一个完整的集中式日志系统，需要包含以下几个主要特点：
收集－能够采集多种来源的日志数据
传输－能够稳定的把日志数据传输到中央系统
存储－如何存储日志数据
分析－可以支持 UI 分析
警告－能够提供错误报告，监控机制 ELK 提供了一整套解决方案，并且都是开源软件，之间互相配合使用，完美衔接，高效的满足了很多场合的应用。目前主流的一种日志系统。

2. ELK 简介

ELK 是三个开源软件的缩写，分别表示：Elasticsearch , Logstash, Kibana , 它们都是开源软件。新增了一个 FileBeat，它是一个轻量级的日志收集处理工具(Agent)，Filebeat 占用资源少，适合于在各个服务器上搜集日志后传输给 Logstash，官方也推荐此工具。

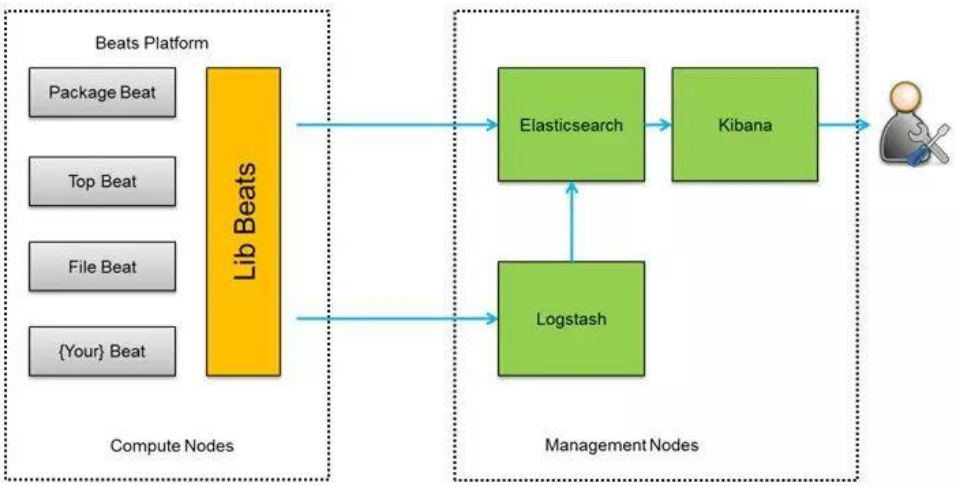
Elasticsearch 是个开源分布式搜索引擎，提供搜集、分析、存储数据三大功能。它的特点有：分布式，零配置，自动发现，索引自动分片，索引副本机制，restful 风格接口，多数据源，自动搜索负载等。

Logstash 主要是用来日志的搜集、分析、过滤日志的工具，支持大量的数据获取方式。一般工作方式为 c/s 架构，client 端安装在需要收集日志的主机上，server 端负责将收到的各节点日志进行过滤、修改等操作在一并发往 elasticsearch 上去。

Kibana 也是一个开源和免费的工具，Kibana 可以为 Logstash 和 ElasticSearch 提供的日志分析友好的 Web 界面，可以帮助汇总、分析和搜索重要数据日志。

Filebeat 隶属于 Beats。目前 Beats 包含四种工具：

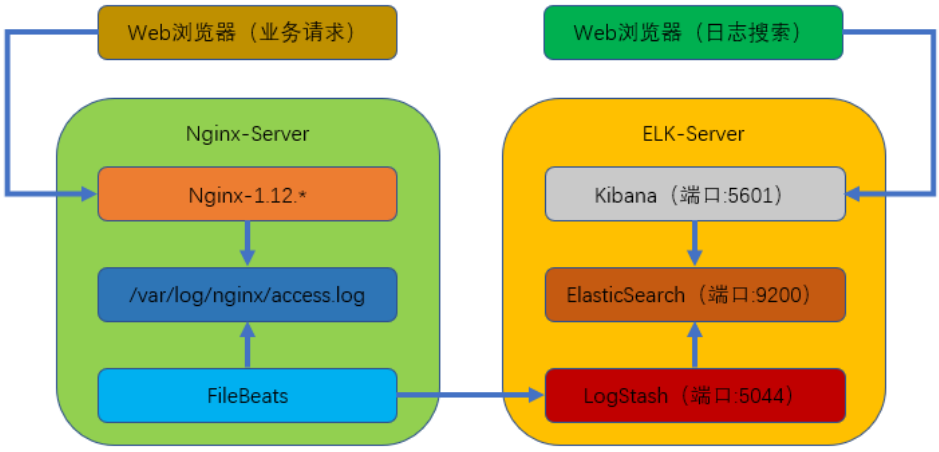
Packetbeat（搜集网络流量数据）
Topbeat（搜集系统、进程和文件系统级别的 CPU 和内存使用情况等数据）
Filebeat（搜集文件数据）
Winlogbeat（搜集 Windows 事件日志数据）



3. 实验部署

本次部署的是 filebeats(客户端)，logstash+elasticsearch+kibana(服务端)组成的架构。

业务请求到达 nginx-server 机器上的 Nginx； Nginx 响应请求，并在 access.log 文件中增加访问记录； FileBeat 搜集新增的日志，通过 LogStash 的 5044 端口上传日志； LogStash 将日志信息通过本机的 9200 端口传入到 ElasticSerach； 搜索日志的用户通过浏览器访问 Kibana，服务器端口是 5601；Kibana 通过 9200 端口访问 ElasticSerach；



实验环境：
本次部署的是单点 ELK 用了两台机器(CentOS-7.5)
ELK 服务端：192.168.88.100
Nginx 客户端：192.168.88.110

1. 准备工作：

配置好网络 yum 源

wget http://mirrors.aliyun.com/repo/Centos-7.repo

wget http://mirrors.aliyun.com/repo/epel-7.repo

关闭防火墙：systemctl stop(disable) firewalld

关闭 SELinux：SELINUX=disabled

2. 下载并安装软件包：

mkdir /elk;cd /elk

wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-6.2.3.tar.gz

wget https://artifacts.elastic.co/downloads/logstash/logstash-6.2.3.tar.gz

wget https://artifacts.elastic.co/downloads/kibana/kibana-6.2.3-linux-x86_64.tar.gz

全部解压缩，并复制到/usr/local/目录下

3. 安装 JDK(java)环境工具：

```
# yum -y install java-1.8*
```

4. 配置 elasticsearch：

1) 新建 elasticsearch 用户并启动(用 elasticsearch 普通用户启动)

```
# useradd elasticsearch
# chown -R elasticsearch.elasticsearch /usr/local/elasticsearch-6.2.3/
# su - elasticsearch
# cd /usr/local/elasticsearch-6.2.3/
# ./bin/elasticsearch -d
```

2) 查看进程是否启动成功（等待一下）

```
# netstat -antp
```

3) 若出现错误可以查看日志

```
# cat /usr/local/elasticsearch-6.2.3/logs/elasticsearch.log
```

4) 测试是否可以正常访问

```
# curl localhost:9200
```

5. 配置 logstash

Logstash 收集 nginx 日志之使用 grok 过滤插件解析日志，grok 作为一个 logstash 的过滤插件，支持根据模式解析文本日志行，拆成字段。

1) logstash 中 grok 的正则匹配

```
vim vendor/bundle/jruby/2.3.0/gems/logstash-patterns-core-4.1.2/patterns/grok-patterns
WZ ([^ ]*)
NGINXACCESS %{IP:remote_ip} \- \- \[%{HTTPDATE:timestamp}\] "%{WORD:method} %{WZ:request}
HTTP/%{NUMBER:httpversion}" %{NUMBER:status} %{NUMBER:bytes} %{QS:referer} %{QS:agent}
 %{QS:xforward}
```

2) 创建 logstash 配置文件

```
vim /usr/local/logstash-6.2.3/default.conf
input {
    beats {
```

```
        port => "5044"
    }
}
#数据过滤
filter {
    grok {
        match => { "message" => "%{NGINXACCESS}" }
    }
    geoip {
        # nginx 客户端 ip
        source => "192.168.88.110"
    }
}
#输出配置为本机的 9200 端口，这是 ElasticSerach 服务的监听端口
output {
    elasticsearch {
        hosts => ["127.0.0.1:9200"]
    }
}
```

3) 进入到/usr/local/logstash-6.2.3 目录下，并执行下列命令

```
后台启动 logstash：nohup bin/logstash -f default.conf &
查看启动日志：tailf nohup.out
查看端口是否启动：netstat -napt|grep 5044
```

6. 配置 kibana

1) 打开 Kibana 配置文件/usr/local/kibana-6.2.3-linux-x86_64/config/kibana.yml，找到下面这行并修改

```
# vim /usr/local/kibana-6.2.3-linux-x86_64/config/kibana.yml
#server.host: "localhost"
修改为
server.host: "192.168.88.100"
这样其他电脑就能用浏览器访问 Kibana 的服务了；
```

2) 进入 Kibana 的目录：cd /usr/local/kibana-6.2.3-linux-x86_64

```
执行启动命令：nohup bin/kibana &
查看启动日志：tail -f nohup.out
查看端口是否启动：netstat -napt|grep 5601
```

3) 测试：

在浏览器访问 192.168.88.100:5601

到此。ELK 部署完成

7. Nginx 客户端配置

1) yum 安装二进制 nginx 软件包

```
# yum -y install nginx
```

2) 下载 filebeat 并解压到/usr/local/

```
# wget https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-6.2.3-linux-x86_64.tar.gz
# tar -xf ./filebeat-6.2.3-linux-x86_64.tar.gz -C /usr/local/
```

3) 打开文件/usr/local/filebeat-6.2.3-linux-x86_64/filebeat.yml，找到如下位置：修改三处

```
enable：false                    #修改为 true
paths：/var/log/*.log            #修改为/var/log/nginx/*.log
#output.elasticsearch:           #将此行注释掉
#hosts: ["localhost:9200"]       #将此行注释掉
output.logstash:                 #取消此行注释
hosts: ["192.168.88.100:5044"]   #取消此行注释并修改 IP 地址为 ELK 服务器地址
```

4) 切换到/usr/local/filebeat-6.2.3-linux-x86_64 目录下

```
# cd /usr/local/filebeat-6.2.3-linux-x86_64
后台启动 filebeat：nohup ./filebeat -e -c filebeat.yml &
查看日志：tailf nohup.out
```

5) 通过浏览器多访问几次 nginx 服务，这样能多制造一些访问日志，访问地址：https://192.168.137.131

6) 访问 Kibana：https://192.168.88.100:5601，点击左上角的 Discover，就可以看到访问日志已经被 ELK 搜集了，然后按照下列步骤完成设置

- 输入 logstash-*，点击”Next step”
- 选择 Time Filter，再点击“Create index pattern”
- 然后可自行创建日志内容查询规则