实训任务六：基于 Apache Log 分析网页业务服务形态

一、目的

1、了解 Apache 日志；

2、实现对 Apache 日志分析；

3、实现网页业务服务状态分析。

二、学时

4 学时

扫码看实训演示

扫码看任务步骤

三、类型

综合型

四、需求

1、硬件

每人配备计算机 1 台，不低于双核 CPU、8G 内存、500GB 硬盘。

每小组配备服务器 1 台。

2、软件

Windows 操作系统，安装 puTTY 管理终端软件，安装 VMware ESXi 控制台软件。

服务器安装 VMware ESXi 7.0。

3、网络

计算机、服务器、虚拟主机使用固定 IP 地址接入局域网，并支持对互联网的访问。

4、工具

无。

五、任务要求

1、完成 Filebeat 的安装与配置；

2、实现 Apache Log 的分析以实现网页业务服务状态分析。

六、考核要求

1、提交《Apache Log 数据分析报告》；

2、提交 Apache Log 可视化分析成果截图/演示视频。

七、任务步骤

本任务需采集 Apache Log 的虚拟机列表如表 6-1 所示。

表 6-1 虚拟机列表

序号	虚拟机名称
1	YWSX-10.10.2.162-CentOS8X64-Cacti

任务 1：配置 Apache

步骤 01：Apache 日志配置

在 Apache 配置文件中可配置日志格式、存储位置等内容，Apache 配置文件默认路径为/etc/httpd/conf/httpd.conf，配置内容如下。

配置文件：

```
LogLevel warn
ErrorLogFormat "[%t] [%l] [pid %P] %F: %E: [client %a] %M"
ErrorLog "logs/error_log"
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\"" combined
CustomLog "|$/usr/local/apache/bin/rotatelogs    /etc/httpd/logs/access_log 86400" combined
```

步骤 02：重载 Apache 配置

参考命令：

```
systemctl reload httpd
```

任务 2：使用 Filebeat 采集 Apache Log 数据

步骤 01：获取 Filebeat

从 https://www.elastic.co/cn/downloads/beats/filebeat 获取安装的 RPM 64-BIT 文件，并上传至 YWSX-10.10.2.162-CentOS8X64-Cacti 虚拟机的/opt 目录。

步骤 02：安装 Filebeat

参考命令：

```
rpm -ivh filebeat-7.8.0-x86_64.rpm
```

步骤 03：配置 Filebeat

Filebeat 配置文件为/etc/filebeat/filebeat.yml。

配置文件：

```
filebeat.inputs:
- type: log
  enabled: true
  paths:
    - /etc/httpd/logs/*
  fields:
    indextype: apachelog-10.10.2.162
  fields_under_root: true

output.Logstash:
  hosts: ["10.10.2.155:5044"]
```

步骤 04：启动 Filebeat

参考命令：

```
filebeat -e -c /etc/filebeat/filebeat.yml
```

任务 3：使用 Logstash 清洗与格式化数据

步骤 01：配置 Logstash

创建 Logstash 配置文件，配置文件路径为/etc/logstash/conf.d/logstash_httpd.conf，配置文件内容如下。

配置文件：

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

filter {
  grok {
    match => { "message" => "%{COMBINEDAPACHELOG}" }
  }
  geoip {
    source => "clientip"
  }
  mutate {
    remove_field => "agent"
    remove_field => "host"
    remove_field => "@version"
    remove_field => "ecs"
    remove_field => "input"
  }
}

output {
  if [indextype] == "apachelog-10.10.2.162" {
    elasticsearch{
      hosts => ["10.10.2.152:9200"]
      index => "apachelog-10.10.2.162-%{+YYYY.MM.dd}"
    }}
}
```

步骤 02：重启 Logstash

参考命令：

```
logstash -f /etc/logstash/conf.d/logstash_winlogbeat.conf
```

任务 4：在 Kibana 上创建分析模型并进行分析

步骤 01：设计分析模型

表 6-2 Syslog 分析模型

序号	分析字段	分析模型	图表类型
1	_id	网站访问请求总次数	指标
2	bytes	发送字节总数量	指标
3	clientip	来源客户端总数量	指标

4	_id+timestamp	网站访问趋势	折线图
5	_id+MONTHDAY	网站每日访问趋势	饼状图
6	-id+request	网站请求次数排行	表格
7	_id+agent	操作系统访问排行	表格
8	_id+response	网站响应状态	饼状图
9	_id+verb	网站请求类型	饼状图
10	_id+geoip.region_name	网站访问区域排行	表格

步骤 02：绘制分析模型

依据表 6-2 分型模型实现图表可视化，如图 6-1 所示。

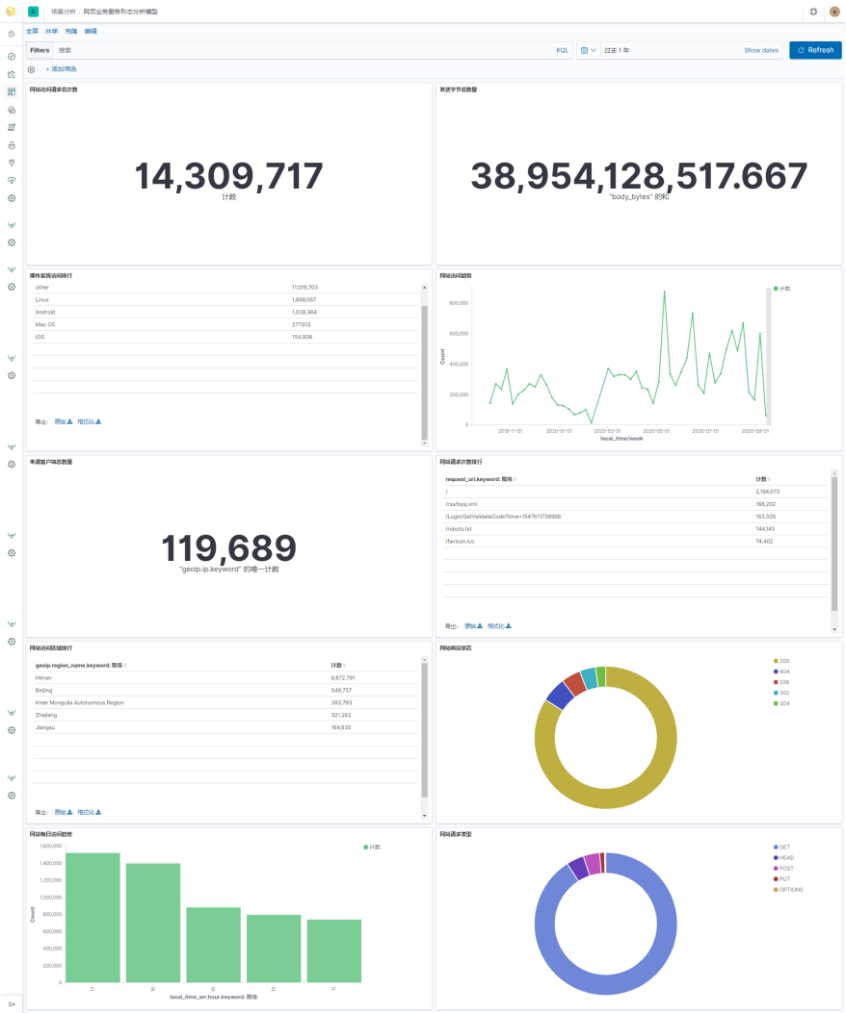

图 6-1 Apache 数据分析

八、实验思考

1、了解 Filebeat。

（1）FileBeat 怎么保持文件的状态？

2、深入理解 Apache Log。

（1）Apache Log 分为几种？有多少种等级？

（2）Apache Log 的格式是什么？

（3）哪些指令可以配置 Apache Log？作用分别是什么？

3、深入了解网页业务服务形态。

（1）哪些指标可以反应出网页业务服务形态？分别反应了什么？

（2）从分析模型中分析出网站存在的不足，并提出解决方法。