云搜索服务日志分析场景实战

1　任务介绍

本次任务，通过对网站 apache 日志的分析，介绍华为云搜索服务集群创建、使用，并向大家展示内置图形化工具 Kibana 的效果。

2　任务执行

2.1　准备环境

2.1.1　创建虚拟私有云

步骤 1 从华为云官网进入虚拟私有云产品页。

步骤 2 点击“立即购买”登录华为云控制台。

步骤 3　选择左上角区域为“华北-北京一”。

步骤 4 在“网络控制台”页面右上角单击“创建虚拟私有云”，已创建过“vpc-css-demo”的可以跳到 2.1.2。

步骤 5 根据界面提示配置虚拟私有云参数后，单击“立即创建”。

- “区域”保持默认为“华北-北京一”。
- “名称”配置为“vpc-css-demo”，也可以按规范命名。
- “网段”保持默认值。
- “标签”保持默认值，不填写。
- “可用分区”保持默认值“可用区 2”。
- “子网名称”配置为“subnet-css-demo”，也可以按规范命名。
- “子网网段”保持默认值。
- “高级配置”选择默认配置。

参数配置完成后，如图所示。

步骤 6 确认安全组 Sys-default 的出入方向都允许所有协议和所有端口访问。

步骤 7 若没有上述规则，请添加。否则，略过此步骤。

2.1.2 购买弹性云服务器

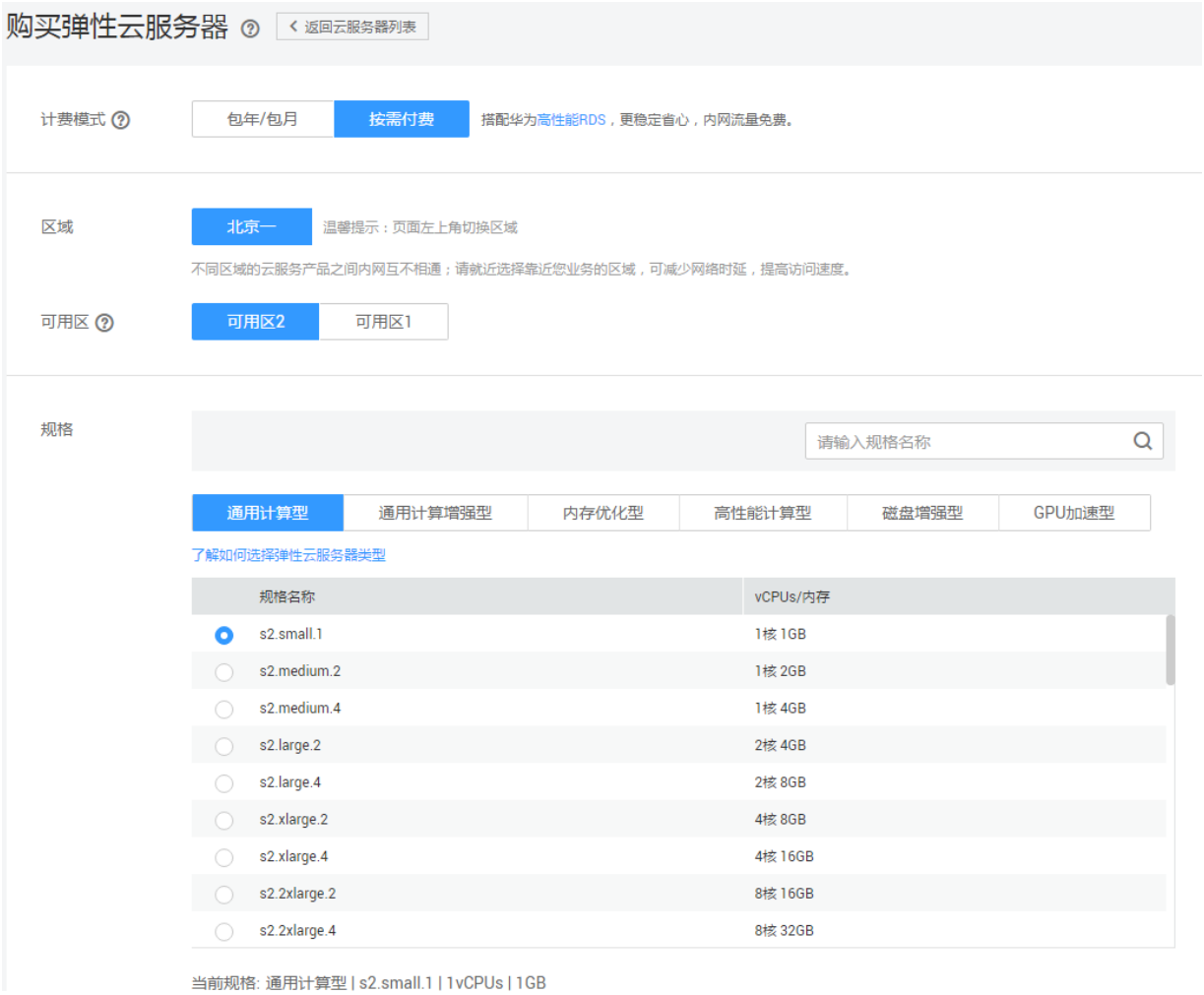
步骤 1 在从华为云官网进入弹性云服务器产品页。

步骤 2 点击“立即购买”进入购买页面。

步骤 3 根据界面提示配置参数后，单击“立即购买”。已购买过可以跳到 2.1.3。

- “计费模式”选择“按需付费”。
- “区域”保持默认值“北京一”。
- “可用区”保持默认值“可用区 2”
- “规格”这里选择“通用计算性 – s2.small.1”。
- “镜像”选择“公共镜像 - EulerOS – EulerOS 2.2 64bit(40GB)”。
- “磁盘 – 系统盘”选择“普通 IO – 50GB”，其他保持默认。
- “自动备份”保持默认不勾选。
- “虚拟私有云”选择我们之前创建的“vpc-css-demo”。
- “安全组”选择“Sys-default”。
- “网卡 – 主网卡”选择“subnet-css-demo”，其他保持默认。
- “弹性 IP”选择“现在购买”。
- “规格”保持默认选择“静态 BGP”。
- “带宽类型”保持默认选择“独享带宽”。
- “计费方式”保持默认选择“按带宽计费”。
- “带宽”选择“5Mbit/s”。
- “登录方式”选择“密码”，并输入登录弹性云服务器的 root 账户密码。请牢记密码，后面会使用。
- “高级配置”保持默认“暂不配置”。
- “云服务器名称”输入“ecs-css-demo”，不勾选允许重名。
- “购买数量”选择“1”。

参数配置后，如下图：


购买弹性云服务器
返回云服务器列表
计费模式
包年/包月
按需付费
搭配华为高性能RDS，更稳定省心，内网流量免费。
区域
北京一
温馨提示：页面左上角切换区域
不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。
可用区
可用区2
可用区1
规格
请输入规格名称
通用计算型
通用计算增强型
内存优化型
高性能计算型
磁盘增强型
GPU加速型
了解如何选择弹性云服务器类型
规格名称
vCPUs/内存
s2.small.1
1核 1GB
s2.medium.2
1核 2GB
s2.medium.4
1核 4GB
s2.large.2
2核 4GB
s2.large.4
2核 8GB
s2.xlarge.2
4核 8GB
s2.xlarge.4
4核 16GB
s2.2xlarge.2
8核 16GB
s2.2xlarge.4
8核 32GB
当前规格: 通用计算型 | s2.small.1 | 1vCPUs | 1GB

镜像 公共镜像 私有镜像 共享镜像 市场镜像

EulerOS EulerOS 2.2 64bit(40GB)

磁盘 云硬盘

系统盘 普通IO 50 GB

增加一块数据盘 您还可以挂载 23 块磁盘（云硬盘）

自动备份 使用自动备份 推荐 建议备份云服务器数据，开通备份会产生额外费用。 1GB数据每月仅需0.268元（购买套餐包，价格更优惠）。

虚拟私有云 vpc-css-demo 查看虚拟私有云

安全组 如何配置安全组？

Sys-default (入方向:TCP/3389, 22 | 出方向: -) 管理安全组

入方向: TCP/3389, 22 | 出方向: -

网卡 主网卡 subnet-css-demo(192.168.0.0/... 自动分配IP地址 查看已使用IP地址

增加一块网卡 您还可以增加 11 块网卡

弹性IP 如有互联网访问需求，请先规划您的弹性IP资源。单击这里查看弹性IP。

现在购买 使用已有 暂不购买

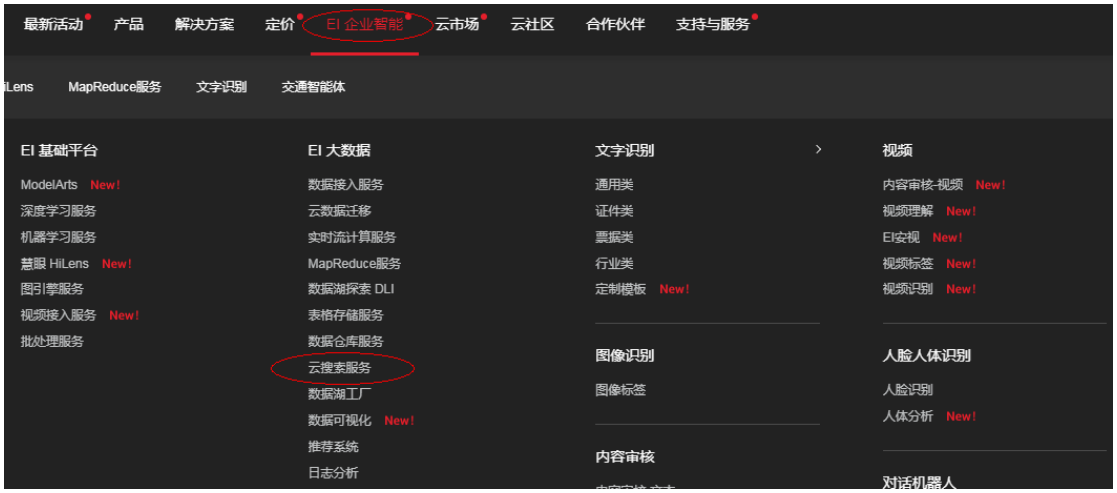
规格 静态BGP 全动态BGP

带宽类型 独享带宽 共享带宽

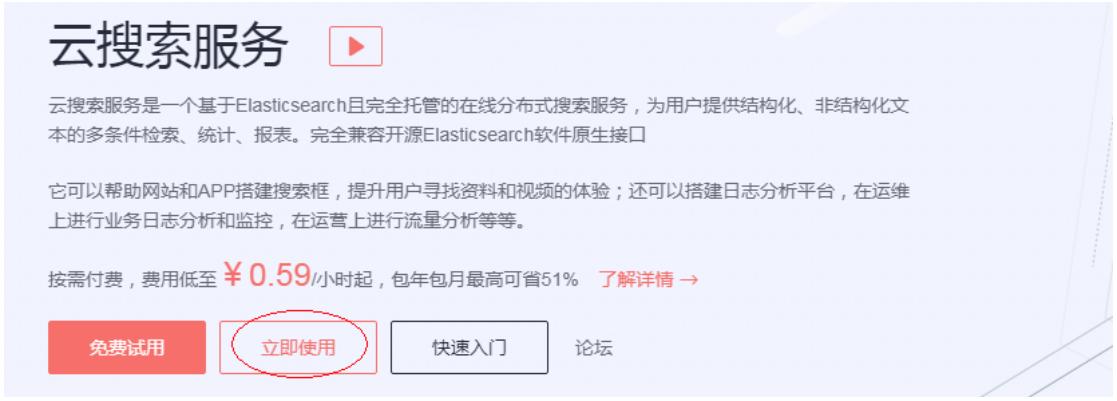
步骤 3 点击“立即购买”，在“规格确认”确认无误后，勾选免责声明，点击“提交申请”，等待 3 分钟左后，弹性云服务器创建成功。

2.1.3 创建云搜索服务集群

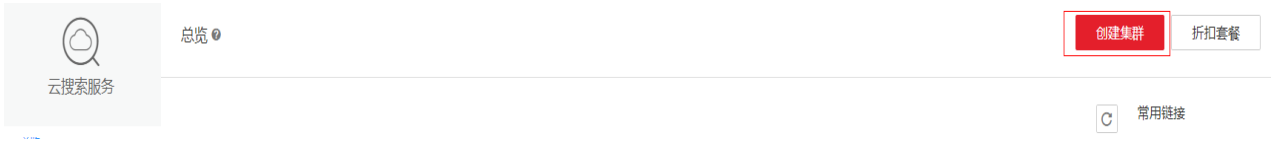
步骤 1 从华为云官网进入云搜索服务产品页。



步骤 2 点击“立即使用”按钮进入控制台页面。



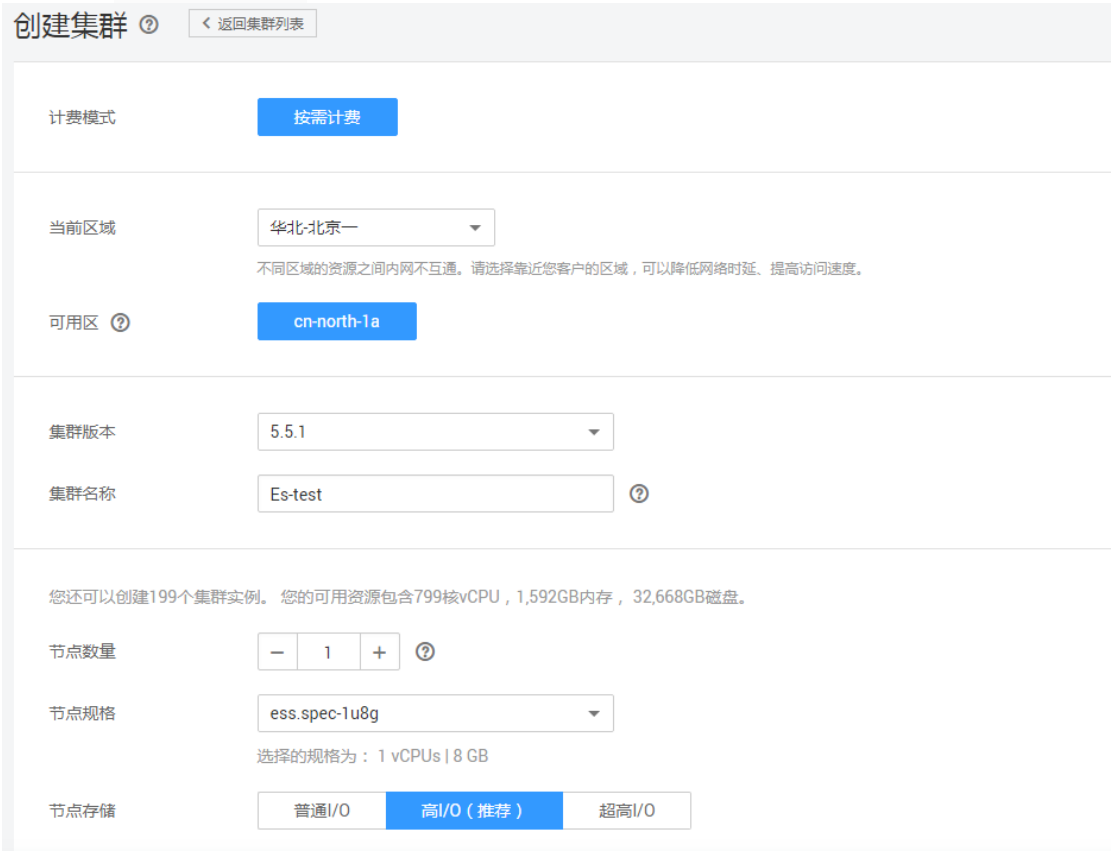
步骤 3 在“云搜索服务”页面右上角单击“创建集群”。之前创建过“Es-test”集群的，可以跳到 2.1.4。

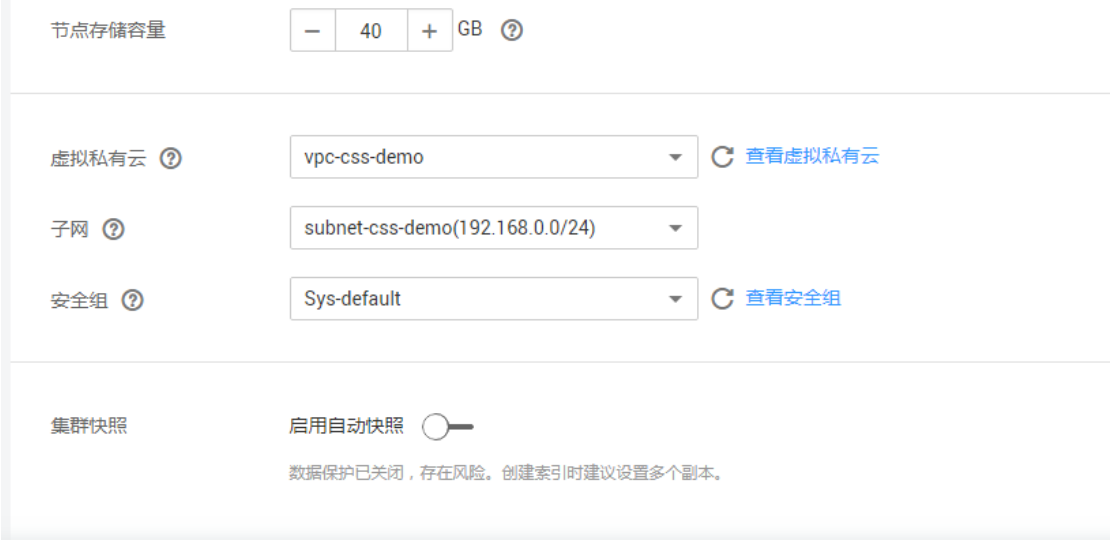


步骤 4 根据界面提示配置参数后，单击“立即申请”。

- “当前区域”从下拉框选择“华北-北京一”
- “集群版本”从下拉框选择，是 Elasticsearch 的软件版本，选择 5.5.1。
- “集群名称”按规范命名配置，可以输入 Es-test。
- “节点数量”自由选择配置，生产集群一般建议 3 个以上。这里选择 1。
- “节点规格”自由选择，生产集群一般建议 ess.spec-2u16g 以上。这里我们选择 ess.spec-1u8g。
- “节点存储”自由选择，这里选择“高 I/O(推荐)”。

- “节点存储容量”可以在允许范围内自由选择。这里输入 40GB。
- “虚拟私有云” VPC 即虚拟私有云，是通过逻辑方式进行网络隔离，提供安全、隔离的网络环境。选择刚才创建的“vpc-css-demo”。
- “子网”：通过子网提供与其他网络隔离的、可以独享的网络资源，以提高网络安全。选择“subnet-css-demo”。
- “安全组”：安全组是一个逻辑上的分组，为同一个 VPC 内具有相同安全保护需求并相互信任的弹性云服务器提供访问策略。选择“Sys-default”。
- “集群快照”：为了增强数据可靠性保障，系统默认开启自动快照功能。这里，我们将其关闭。





步骤 5 在规格确认界面确认完成后，单击“提交申请”开始创建集群。

步骤 6 单击“返回集群列表”，系统将跳转到“集群管理”页面，您创建的集群将展现在集群列表中。且集群状态为“创建中”。

步骤 7 耐心等待十几分钟，创建成功后集群状态会变为“可用”。

2.1.4 安装 java 和 logstash

步骤 1 使用 ssh 工具(比如 putty)登录弹性云服务器，连接 IP 为弹性云服务器列表中的弹性 IP。

步骤 2 在弹性云服务器 ssh 命令行，运行下列命令。安装 java 环境，java -version 测试是否安装成功。输出如下显示安装成功。

```
[root@ecs-css-demo ~]# /usr/bin/yum install java
[root@ecs-css-demo ~]# java -version
openjdk version "1.8.0_171"
OpenJDK Runtime Environment (build 1.8.0_171-b10)
OpenJDK 64-Bit Server VM (build 25.171-b10, mixed mode)
```

步骤 3 下载软件包 css_apache_demo.tar.gz 到本地，并通过 sftp 工具(比如 winscp)上传到弹性云服务器某个目录，这里以/root 目录为例。

步骤 4 在弹性云服务器 ssh 命令行，运行下列命令。解压软件包 css_apache_demo.tar.gz，并进入 css_apache_demo 目录，得到目录 demo_config、demo_logs 和压缩包 logstash-5.5.1.tar.gz。

```
[root@ecs-css-demo ~]# tar -zxf css_apache_demo.tar.gz
[root@ecs-css-demo ~]# cd /root/css_apache_demo/;ll
total 91164
drwx------ 2 3397668  1049089        4096 Jul 19 13:44 demo_config
drwx------ 2 3397668  1049089        4096 Jul 19 14:00 demo_logs
-rw------- 1 3397668  1049089       93242602 Jul 17 10:14 logstash-5.5.1.tar.gz
```

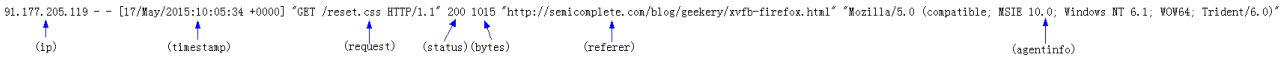
步骤 5 在弹性云服务器 ssh 命令行，运行下列命令。解压 logstash-5.5.1.tar.gz 得到目录 logstash-5.5.1，并更改用户组为 root。

```
[root@ecs-css-demo css_apache_demo]# tar -zxf logstash-5.5.1.tar.gz; chown -R root.root /root/css_apache_demo
[root@ecs-css-demo css_apache_demo]# ll
total 91164
drwx------ 2 root root        4096 Jul 19 13:44 demo_config
drwx------ 2 root root        4096 Jul 19 14:00 demo_logs
drwx------ 11 root root        4096 Jul 19 10:39 logstash-5.5.1
-rw-r--r--  1 root root 93242602 Jul 17 10:14 logstash-5.5.1.tar.gz
```

2.2 准备数据

2.2.1 理解数据

解压 css_apache_demo.tar.gz 后，在 demo_logs/apache_logs 文件中，存放的是某网站产生的 apache 格式日志。一条日志中，包含 ip 信息、timestamp 信息、request 信息、status 返回码信息、bytes 请求流量信息、referer 信息、agentinfo 访问客户端信息。如下图：



83.149.9.216 - - [17/May/2015:10:05:03 +0000] "GET /presentations/logstash-monitorama-2013/images/kibana-search.png HTTP/1.1" 200 203023 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:43 +0000] "GET /presentations/logstash-monitorama-2013/images/kibana-dashboard3.png HTTP/1.1" 200 171717 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:47 +0000] "GET /presentations/logstash-monitorama-2013/plugin/highlight/highlight.js HTTP/1.1" 200 26185 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:12 +0000] "GET /presentations/logstash-monitorama-2013/plugin/zoom-js/zoom.js HTTP/1.1" 200 7697 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:07 +0000] "GET /presentations/logstash-monitorama-2013/plugin/notes/notes.js HTTP/1.1" 200 2892 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:34 +0000] "GET /presentations/logstash-monitorama-2013/images/sad-medic.png HTTP/1.1" 200 430406 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:57 +0000] "GET /presentations/logstash-monitorama-2013/css/fonts/Roboto-Bold.ttf HTTP/1.1" 200 38720 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:50 +0000] "GET /presentations/logstash-monitorama-2013/css/fonts/Roboto-Regular.ttf HTTP/1.1" 200 41820 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:24 +0000] "GET /presentations/logstash-monitorama-2013/images/frontend-response-codes.png HTTP/1.1" 200 52878 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"
83.149.9.216 - - [17/May/2015:10:05:50 +0000] "GET /presentations/logstash-monitorama-2013/images/kibana-dashboard.png HTTP/1.1" 200 321631 "http://semicomplete.com/presentations/logstash-monitorama-2013/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.77 Safari/537.36"

2.2.2 导入数据

步骤 1 在弹性云服务器 ssh 命令行，运行下列命令。输入编辑 demo_config/apache_logstash.conf 文件，修改第 3 行日志文件路径、第 37 行 Elasticsearch 地址(云搜索服务集群列表中的内网访问地址)、 第 39 行 template 路径。

```
[root@ecs-css-demo css_apache_demo]# vi /root/css_apache_demo/demo_config/apache_logstash.conf
```

修改后如图：

```
1 input {
2   file {
3     path => "/root/css_apache_demo/demo_logs/apache_logs"
4     start_position => "beginning"
5   }
6 }
```

```
elasticsearch {
  hosts => "192.168.0.3:9200"
  index => "apache_elastic_example"
  template => "/root/css_apache_demo/demo_config/apache_template.json"
  template_name => "apache_elastic_example"
  template_overwrite => true
}
```

步骤 2 在弹性云服务器 ssh 命令行，运行下列命令。启动 logstash 开始导入数据，至小圆点不再输出后，数据导入完成。但 logstash 仍然运行监听日志文件。

```
[root@ecs-css-demo css_apache_demo]# /root/css_apache_demo/logstash-5.5.1/bin/logstash -f /root/css_apache_demo/demo_config/apache_logstash.conf
```

如下图：

```
[2018-07-19T15:25:13,711][INFO ][logstash.agent           ] Successfully started Logstash API endpoint {:port=>9601}
```

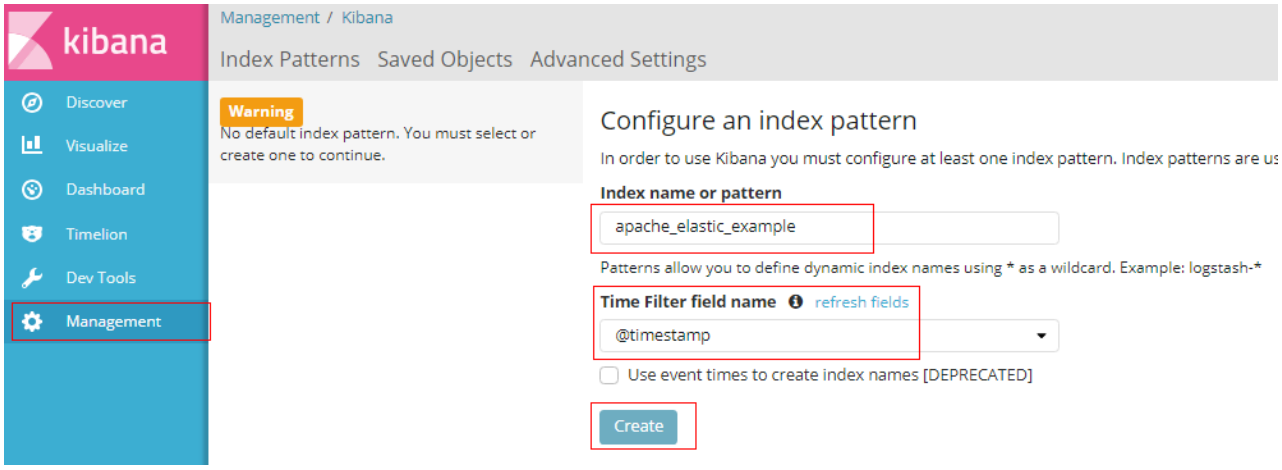
2.3 搜索分析

2.3.1 日志搜索

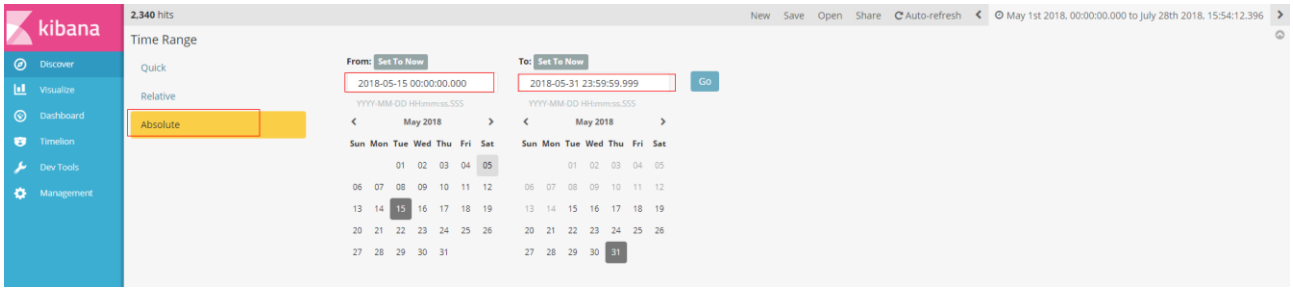
步骤 1 从云搜索服务控制台=>集群列表的“Kibana”链接进入可视化界面。

集群名称	集群状态	任务状态	搜索引擎版本	创建时间	内网访问地址	公网访问地址	操作
Es-apache_log_demo	可用	--	5.5.1	2018/07/19 10:35:03 GMT...	192.168.0.112:9200	--	Kibana 更改规格 更多

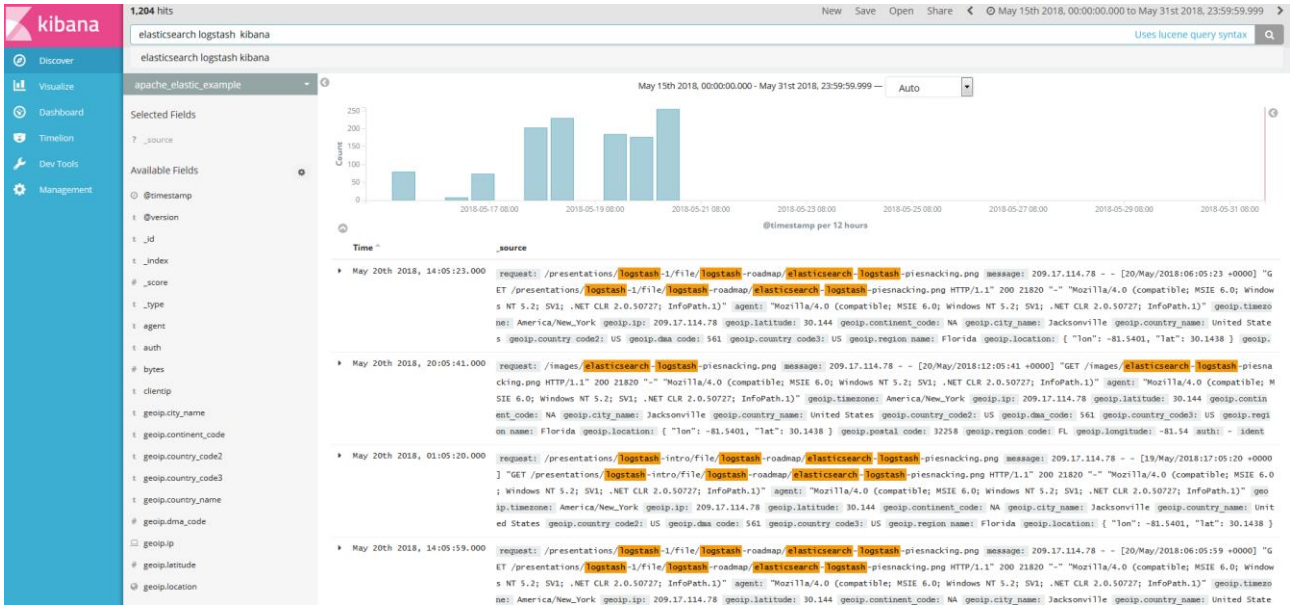
步骤 2 在可视化界面“Management”=> “Index Patterns”页签，输入“apache_elastic_example”作为 index name，选择@timestamp 作为刷新字段，点击 create 按钮创建 index patten。



步骤 3 可视化界面“Discover”页签，点击右上角的时间窗口选择器，将时间窗口定义在“2018-05-15”到“2018-05-31”。

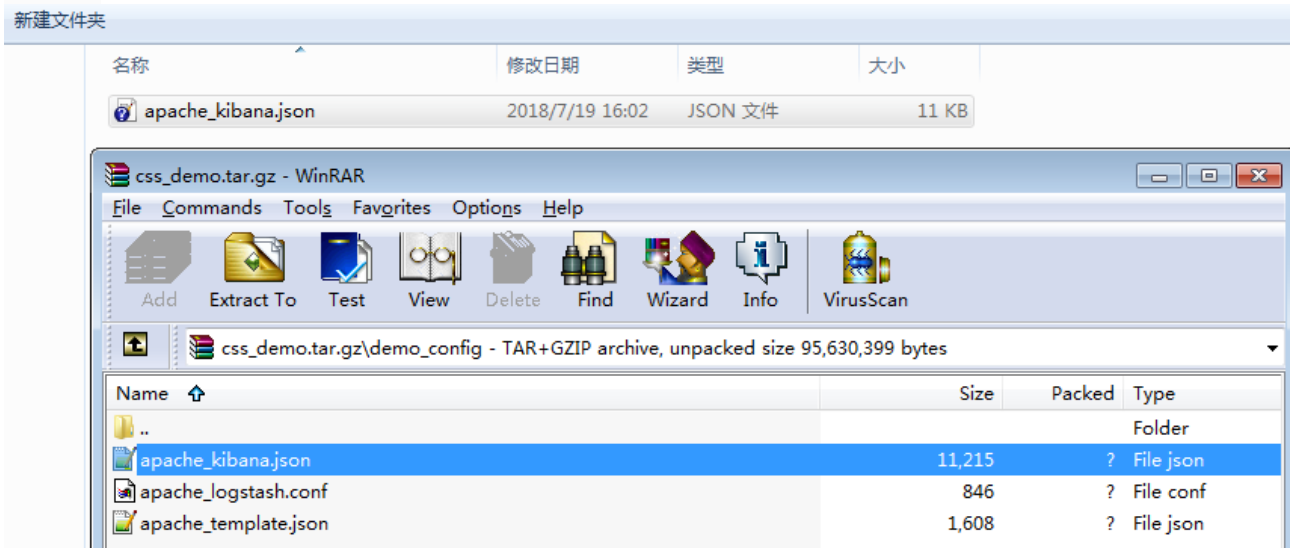


步骤 4 现在你可以像普通搜索引擎一样进行搜索。比如：在搜索框中输入搜索字符串“elasticsearch logstash kibana”，在任何字段信息中，只要任意有匹配的日志信息都找出来了。这样，就可以看到，有没有用户在访问和“elasticsearch logstash kibana”相关的日志信息。



2.3.2 日志分析

步骤 1 在本机上，从软件包 css_apache_demo.tar.gz 中解压 apache_kibana.json 文件到本地。

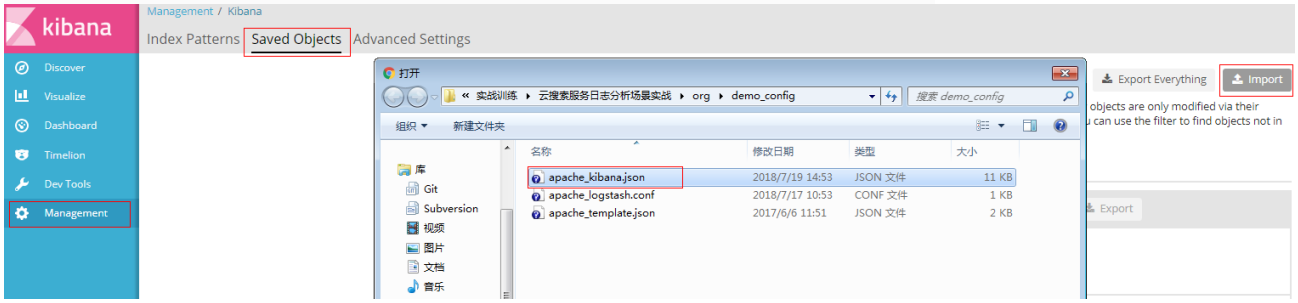


步骤 2 在本机用文本编辑器打开 apache_kibana.json 文件，替换 37 行“UserName”为自己的华为云用户名，保存。

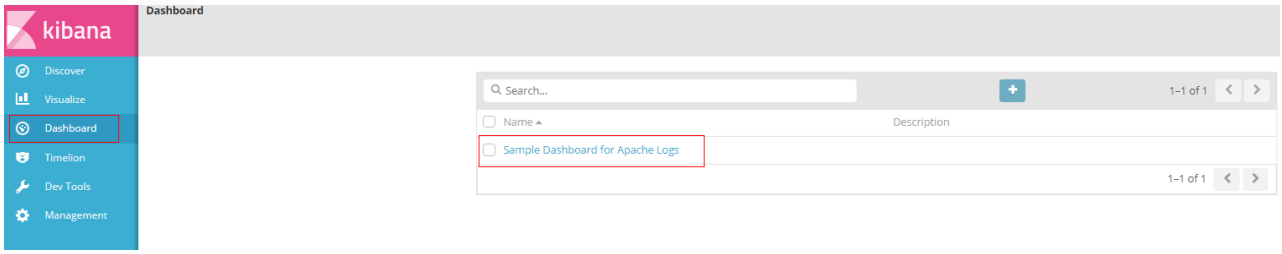
apache_kibana.json

```
},
{
  "_id": "Apache-Dashboard",
  "_type": "visualization",
  "_source": {
    "title": "Overview",
    "visState": "{\"type\":\"markdown\",\"params\":{\"markdown\":\"####
Apache日志可视化分析仪表盘样例_UserName\\n在网站的运营过程中，网站访问者的各种信息对于网站运营者来说，是非常有价值的信息，而网站日志则是这些信息的来源，其记录着web服务器接收处理请求等各种原始信息。\\n\\n本实战场景，以一个网站的apache日志作为举例，体验日志可视化分析场景中[云搜索服务](https://www.huaweicloud.com/product/es.html)的应用。\\n\\n**分析包括:**\\n* 网站独立访问者的数量 \\n* 访问热度地图\\n*
不同国家的访问设备比例关系 \\n*
不同国家的访问操作系统比例关系\\n\\n\"},\"aggs\":[],\"listeners\":{}}",
```

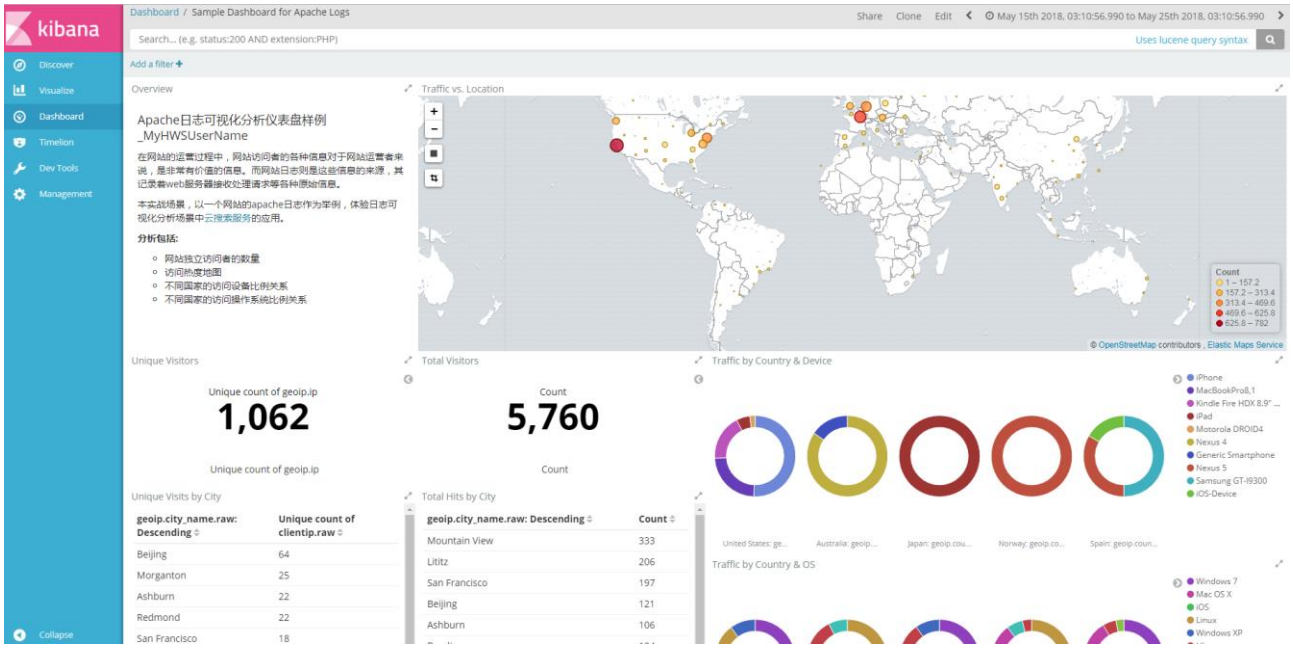
步骤 3 在可视化界面“Management – Saved Objects”页签，点击 import 按钮，选择 apache_kibana.json 文件，并点击“Yes, overwrite all” 导入。



步骤 4 在可视化界面“Dashboard” ，点击“Sample Dashboard for Apache Logs”查看分析图表。



如图：



至此，关于网站的访问情况，已经从多个维度，图形化展现了分析结果。在图形化界面创建这里，我们直接使用了模板导入，您也可以在了解了 Kibana 后，采用托拉拽的方式，自己创建适合自己的图形化界面。

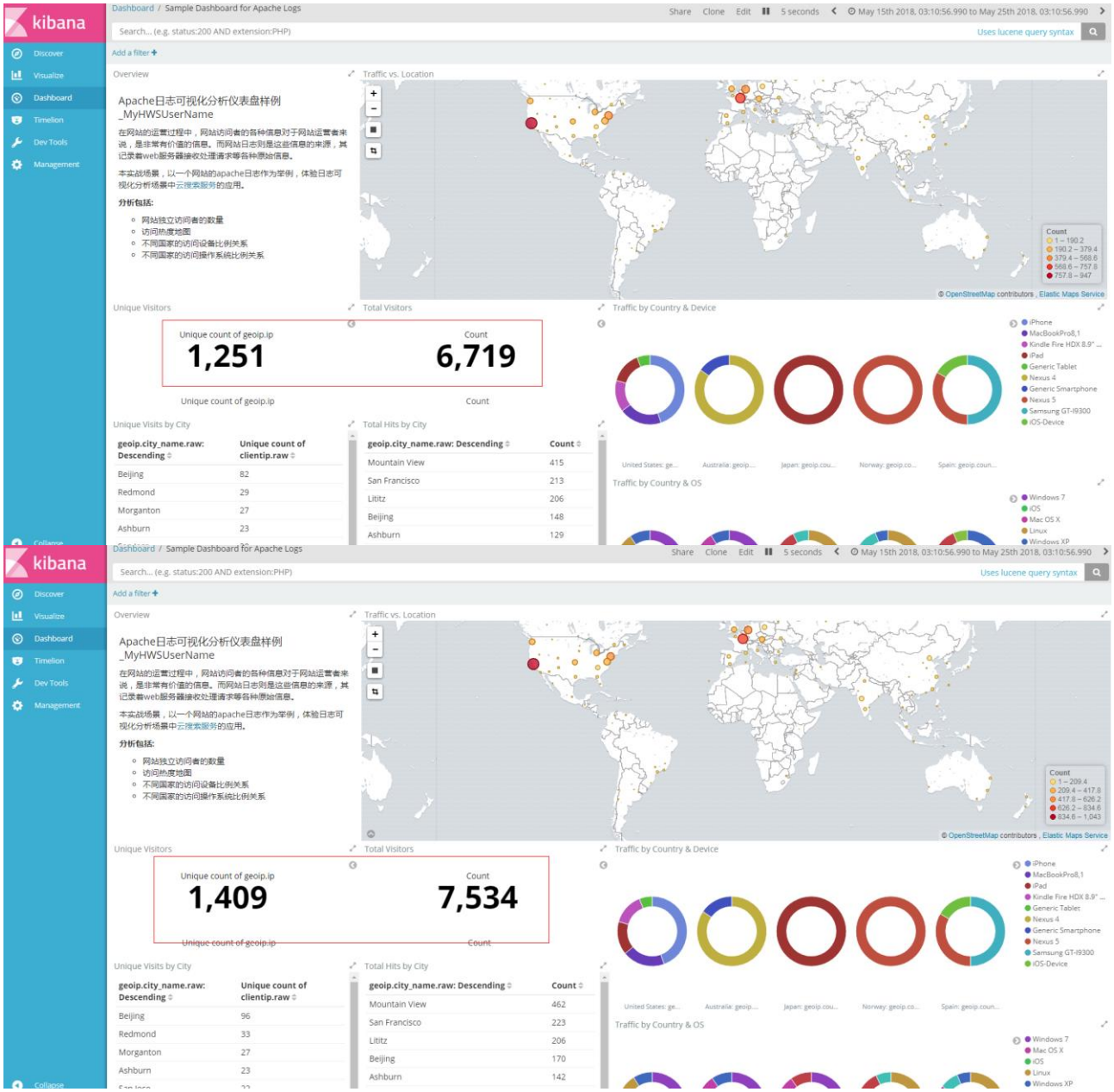
2.3.3 近实时分析

步骤 1 点击 Dashboard 界面右上角的时间窗口，单击 Auto-refresh，并选择刷新间隔(5s)，开启自动刷新功能。

步骤 2 ssh 登录一个新的云服务器命令行，运行下面命令，模拟写入增量日志到 demo_logs/apache_logs 文件。

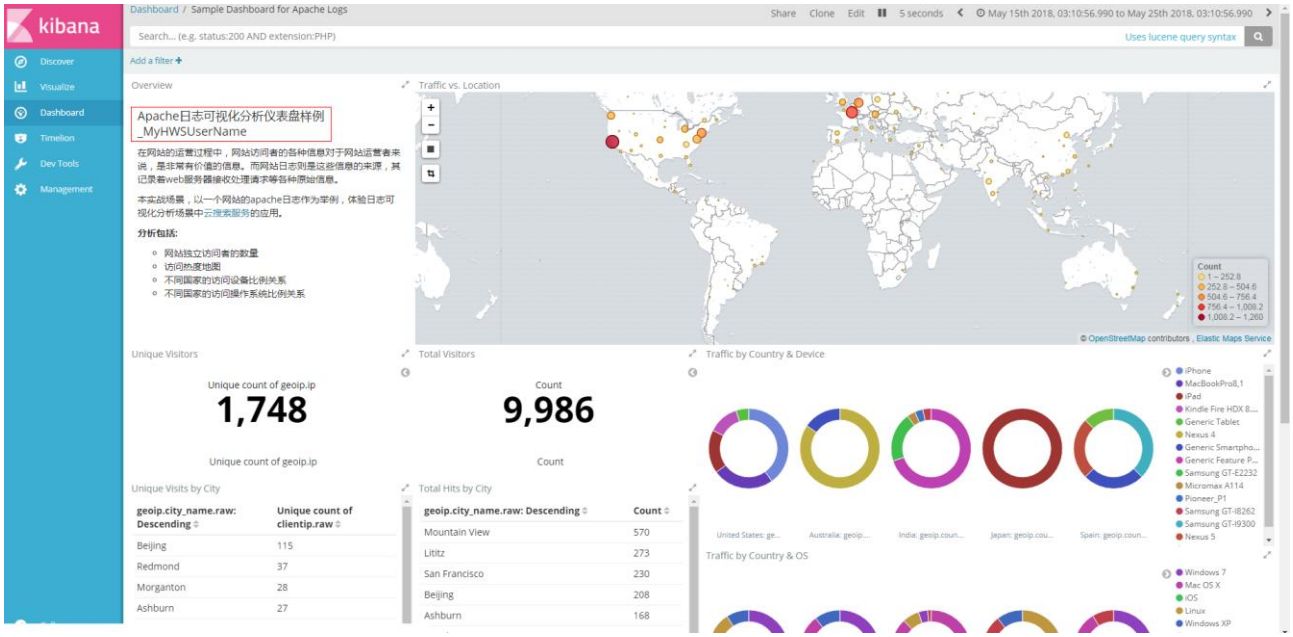
```
[root@ecs-css-demo css_apache_demo]# while read line; do echo $line >>
/root/css_apache_demo/demo_logs/apache_logs; usleep 20000;done <
/root/css_apache_demo/demo_logs/delta_apache_logs
```

回到可视化界面，可以看到可视化界面的数据开始动态刷新：



3 任务打卡

在 kibana 图形化页面，截图带有自己华为云用户名的分析结果，完成打卡。


kibana
Dashboard / Sample Dashboard for Apache Logs
Share
Clone
Edit
5 seconds
May 15th 2018, 03:10:56.990 to May 25th 2018, 03:10:56.990
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Overview
Apache日志可视化分析仪表盘样例
_MyHWSUserName
在网站的运营过程中，网站访问者的各种信息对于网站运营者来说，是非常有价值的信息。而网站日志则是这些信息的来源，其记录着web服务器接收处理请求等各种原始信息。
本实战场景，以一个网站的apache日志作为举例，体验日志可视化分析场景中云搜索服务的应用。
分析包括:
网站独立访问者的数量
访问热度地图
不同国家的访问设备比例关系
不同国家的访问操作系统比例关系
Traffic vs. Location
Count
1 – 252.8
252.8 – 504.6
504.6 – 756.4
756.4 – 1,008.2
1,008.2 – 1,260
© OpenStreetMap contributors , Elastic Maps Service
Unique Visitors
Unique count of geoip.ip
1,748
Unique count of geoip.ip
Total Visitors
Count
9,986
Count
Traffic by Country & Device
iPhone
MacBookPro8,1
iPad
Kindle Fire HDX 8....
Generic Tablet
Nexus 4
Generic Smartpho...
Generic Feature P...
Samsung GT-E2232
Micromax A114
Pioneer_P1
Samsung GT-I8262
Samsung GT-I9300
Nexus 5
United States: ge...
Australia: geoip....
India: geoip.coun...
Japan: geoip.cou...
Spain: geoip.coun...
Unique Visits by City
geoip.city_name.raw: Descending
Unique count of clientip.raw
Beijing 115
Redmond 37
Morganton 28
Ashburn 27
Total Hits by City
geoip.city_name.raw: Descending
Count
Mountain View 570
Lititz 273
San Francisco 230
Beijing 208
Ashburn 168
Traffic by Country & OS
Windows 7
Mac OS X
iOS
Linux
Windows XP