海量日志实时分析

日志搜索引擎

DTCC

2015中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2015

大数据技术探索和价值发现

IT168.com ChinaUnix ITPUB

提纲

- 日志的应用场景
- 过去的做法
- 现在的做法
- 日志搜索引擎
- 日志易产品架构

一条 Apache Access 日志

- 180.150.189.243 - - [15/Apr/2015:00:27:19 +0800] “POST /report HTTP/1.1” 200 21 “https://rizhiyi.com/search/” “Mozilla/5.0 (Windows NT 6.1; WOW64; rv:37.0) Gecko/20100101 Firefox/37.0” “10.10.33.174” 0.005 0.001
- 字段：
  - Client IP: 180.150.189.243
  - Timestamp: 15/Apr/2015:00:27:19 +0800
  - Method: POST
  - URI: /report
  - Version: HTTP/1.1
  - Status: 200
  - Bytes: 21
  - Referrer: https://rizhiyi.com/search/
  - User Agent: Mozilla/5.0 (Windows NT 6.1; WOW64; rv:37.0) Gecko/20100101 Firefox/37.0
  - X-Forward: 10.10.33.174
  - Request_time: 0.005
  - Upstream_request_time:0.001

日志：时间序列机器数据

- IT 系统信息
  - 操作系统
  - 应用软件
- 用户信息
  - 用户行为
- 各种传感器信息
- 日志反映的是事实数据
  - 深度解析LinkedIn大数据平台（http://www.csdn.net/article/2014-07-23/2820811/1）

应用场景

- 运维可用性监控
- 应用性能监控
  - Application Performance Monitoring (APM)
- 安全审计
  - Security Information Event Management (SIEM)
- 用户数据统计分析
- 物联网
  - 智能家电
    - Nest Lab 采集的智能恒温器数据使用 Splunk 来分析
  - 车联网

过去的做法

- 日志没有集中处理
  - 登陆每一台服务器，使用脚本命令或程序查看
- 日志被删除
  - 磁盘满了删日志
  - 黑客删除日志，抹除入侵痕迹
- 日志只做事后追查
  - 没有实时监控、分析
- 使用数据库存储日志
  - 无法适应TB级海量日志
  - 数据库的schema无法适应千变万化的日志格式
  - 无法提供全文检索
- Complex Event Processing (CEP)
  - 难以处理大数据量

现在的做法

- Hadoop
  - 批处理，不够及时
  - 查询慢
  - 可作基于日志的用户数据离线挖掘，无法做 OLAP (On Line Analytic Processing)
- Storm
  - 历史久，停止开发
  - 任务调度差
- Spark
  - 生态圈完整
  - DataBricks 专门支持
- Storm vs. Spark Streaming
  - Storm 是真正的流式处理，Spark Streaming 是 mini-batch
  - Exactly Once vs. At Least Once
  - 延时与吞吐率的取舍
- Hadoop/Storm/Spark都只是一个开发框架，不是拿来即用的产品

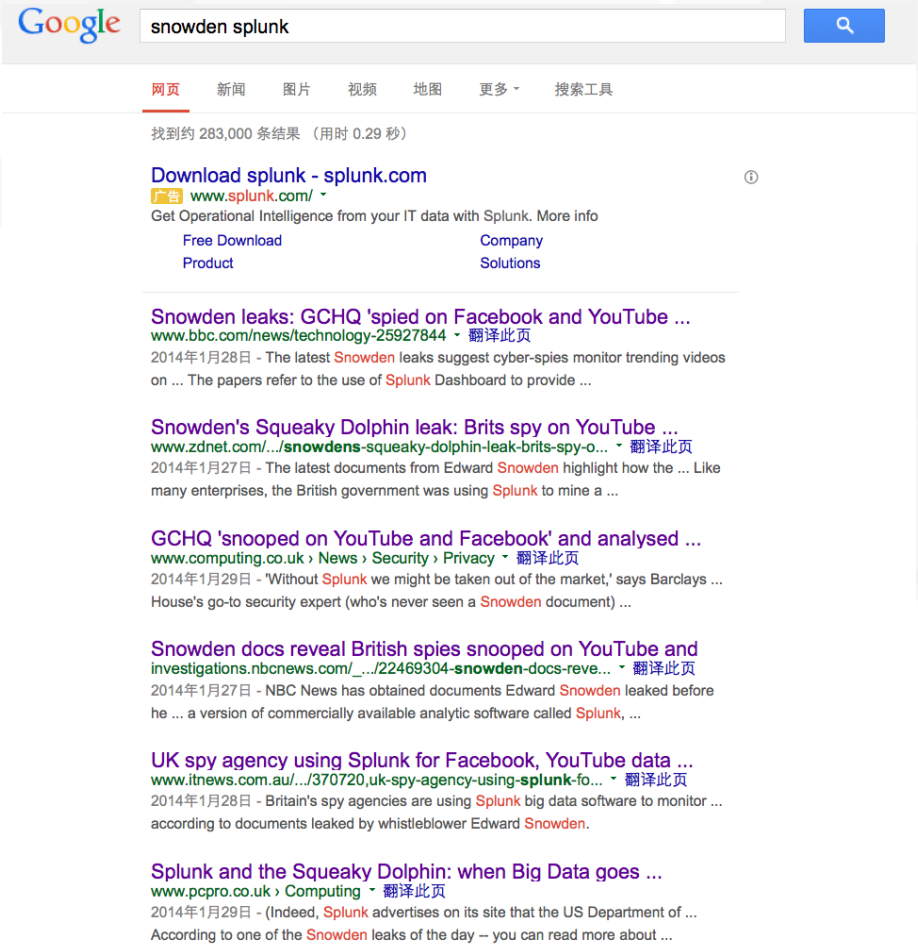
对日志准实时搜索、分析

- 日志准实时搜索引擎
- Splunk
- ELK (Elasticsearch/Logstash/Kibana)

Splunk

- 首创用准实时搜索的方法来分析日志
- 功能非常丰富
  - Search Processing Language
    - 类似 Linux 命令，支持管道，子查询等功能
- Gartner report (2014/7/18)
  - Splunk在日志检索时抽取日志的关键字段，检索速度慢
  - Splunk按每天处理的日志量收费，价格较贵

Splunk 与 棱镜门


Google
snowden splunk
网页 新闻 图片 视频 地图 更多 搜索工具
找到约 283,000 条结果 （用时 0.29 秒）
Download splunk - splunk.com
广告 www.splunk.com/
Get Operational Intelligence from your IT data with Splunk. More info
Free Download
Company
Product
Solutions
Snowden leaks: GCHQ 'spied on Facebook and YouTube ...
www.bbc.com/news/technology-25927844 翻译此页
2014年1月28日 - The latest Snowden leaks suggest cyber-spies monitor trending videos on ... The papers refer to the use of Splunk Dashboard to provide ...
Snowden's Squeaky Dolphin leak: Brits spy on YouTube ...
www.zdnet.com/.../snowdens-squeaky-dolphin-leak-brits-spy-o... 翻译此页
2014年1月27日 - The latest documents from Edward Snowden highlight how the ... Like many enterprises, the British government was using Splunk to mine a ...
GCHQ 'snooped on YouTube and Facebook' and analysed ...
www.computing.co.uk › News › Security › Privacy 翻译此页
2014年1月29日 - 'Without Splunk we might be taken out of the market,' says Barclays ... House's go-to security expert (who's never seen a Snowden document) ...
Snowden docs reveal British spies snooped on YouTube and
investigations.nbcnews.com/_.../22469304-snowden-docs-reve... 翻译此页
2014年1月27日 - NBC News has obtained documents Edward Snowden leaked before he ... a version of commercially available analytic software called Splunk, ...
UK spy agency using Splunk for Facebook, YouTube data ...
www.itnews.com.au/.../370720,uk-spy-agency-using-splunk-fo... 翻译此页
2014年1月28日 - Britain's spy agencies are using Splunk big data software to monitor ... according to documents leaked by whistleblower Edward Snowden.
Splunk and the Squeaky Dolphin: when Big Data goes ...
www.pcpro.co.uk › Computing 翻译此页
2014年1月29日 - (Indeed, Splunk advertises on its site that the US Department of ... According to one of the Snowden leaks of the day -- you can read more about ...

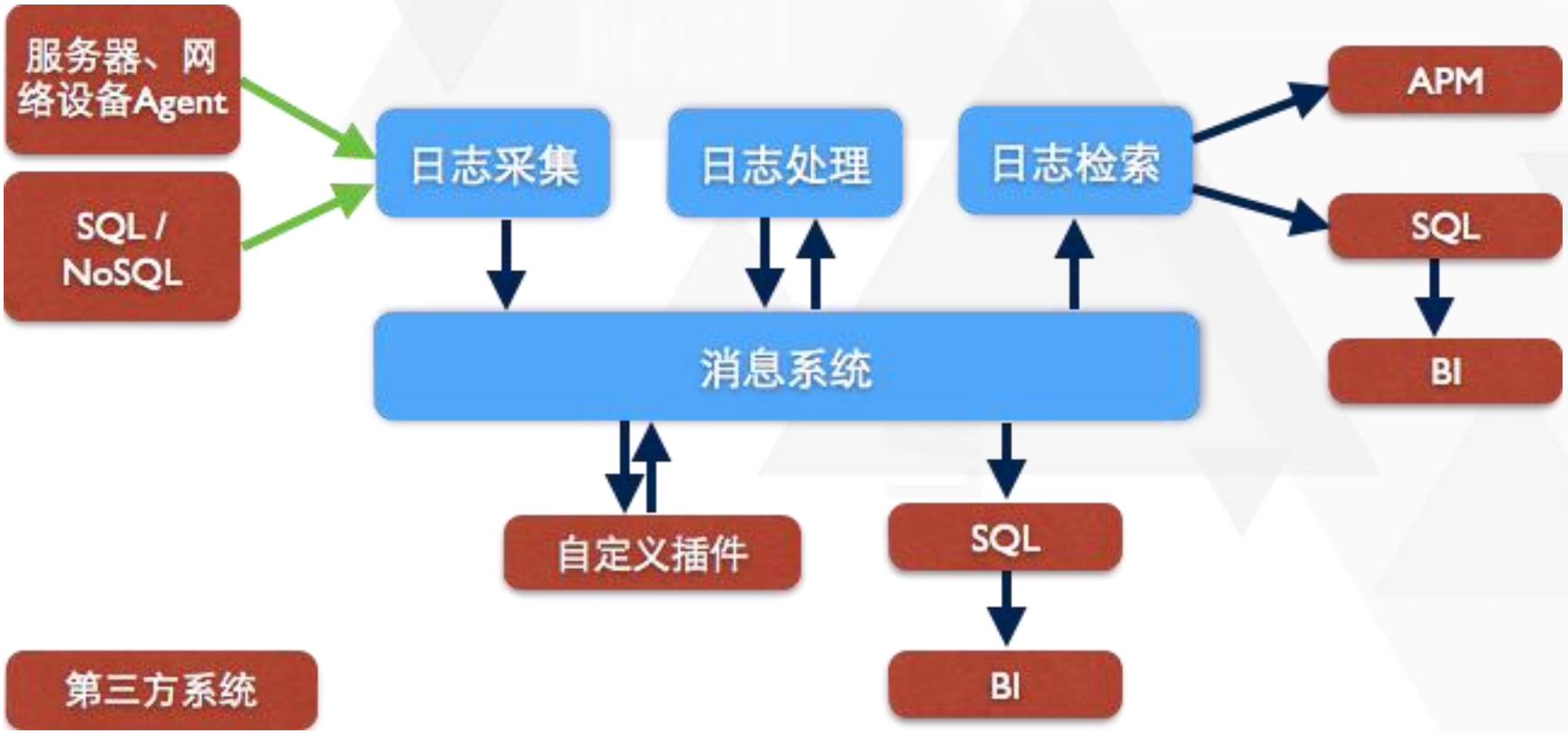
ELK

- 三个独立的开源套件
  - 一些著名互联网公司对 ELK 做二次开发，使用过百台的 ELK 集群分析日志
- 存在问题
  - 运维管理不方便，三个独立的系统，没有统一的部署、管理工具
  - 没有告警功能
  - 没有用户认证及权限管理
  - 统计、分析功能有限

日志易

- 日志搜索分析平台
  - rizhiyi.com
- 企业部署版
- SaaS 版
  - 每天500MB日志处理免费

日志易架构


服务器、网络设备Agent
SQL / NoSQL
日志采集
日志处理
日志检索
APM
SQL
BI
消息系统
自定义插件
SQL
BI
第三方系统

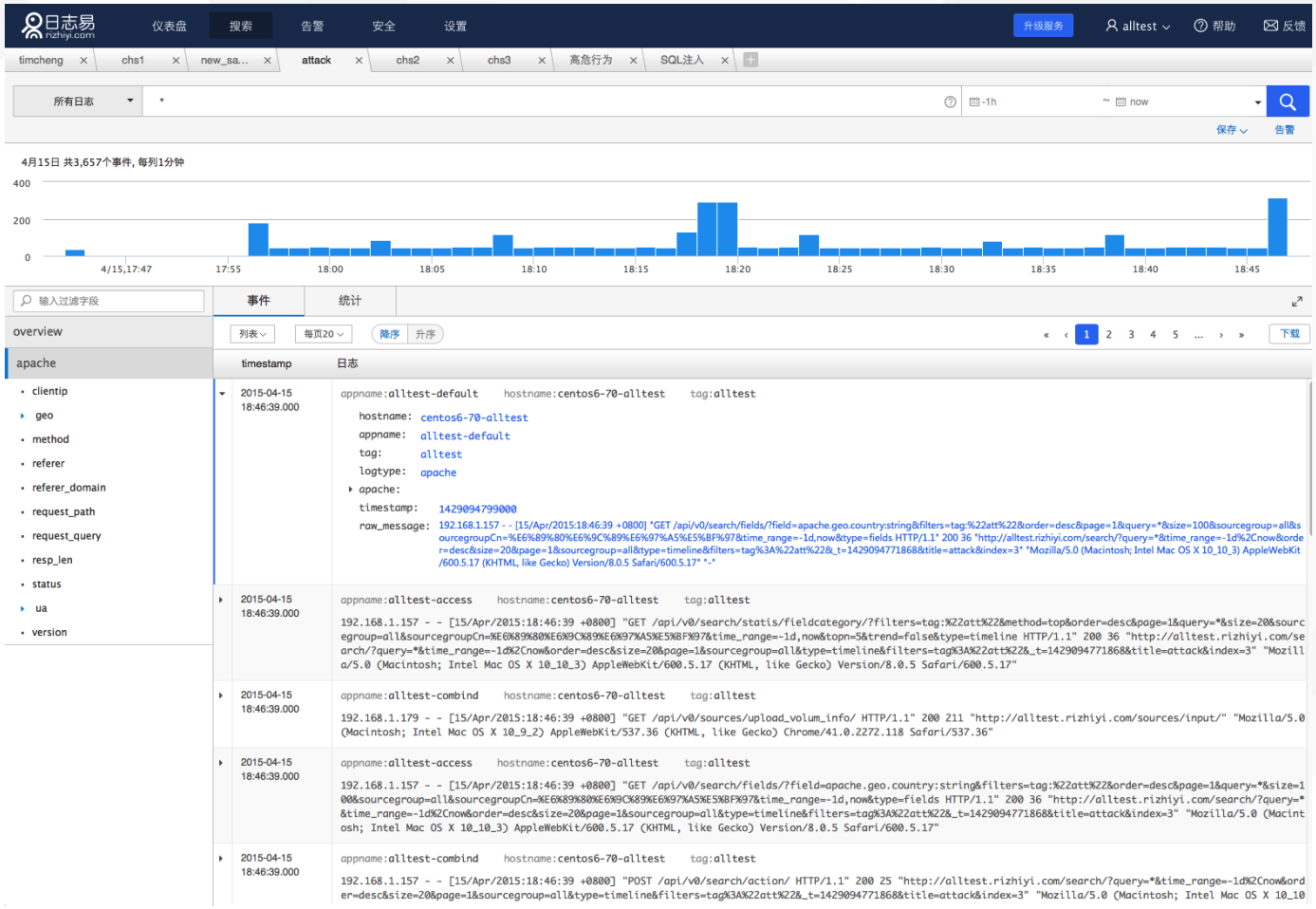
日志易功能

- 搜索
- 告警
- 统计
  - 事务关联
- 配置解析规则，识别任何日志
- 安全攻击自动识别
- 开放API，对接第三方系统
- 高性能、可扩展分布式架构
  - 10万 EPS (Event Per Second), 每天TB级日志

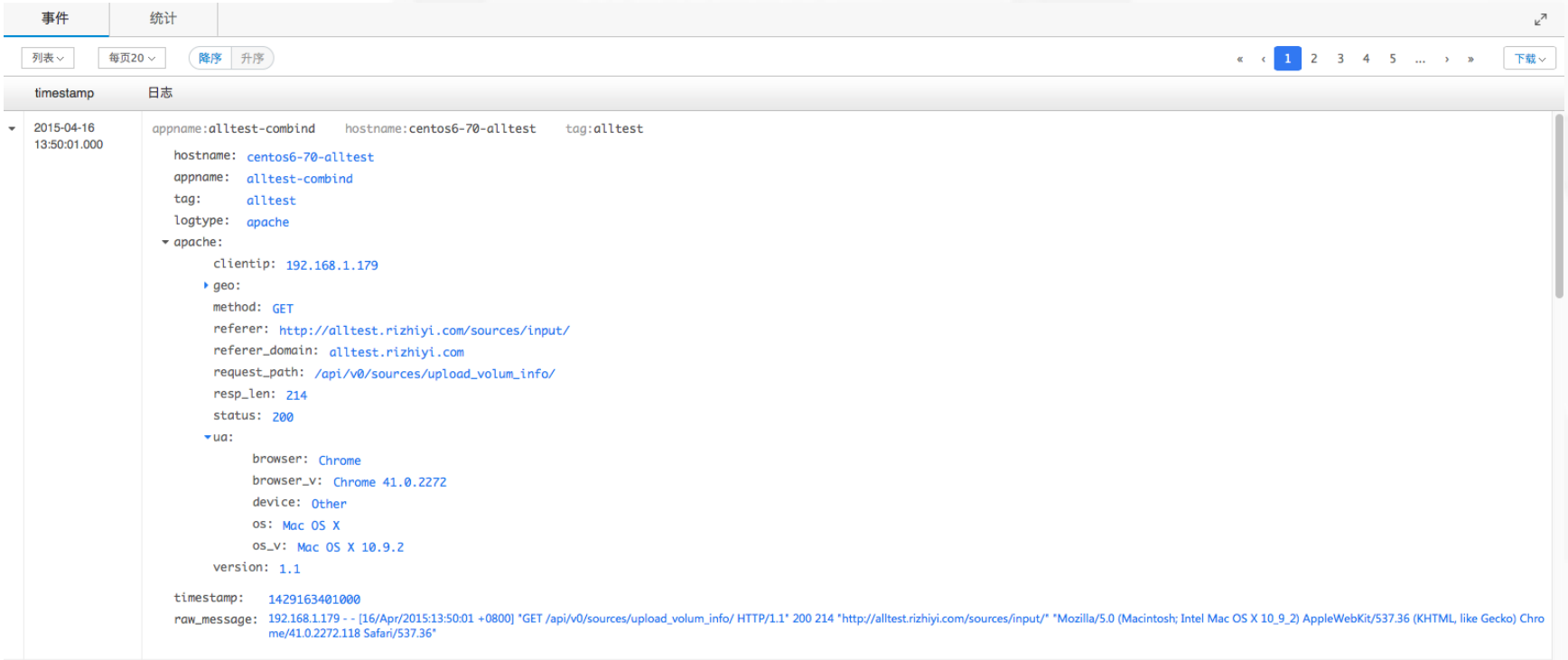
日志易 vs. Splunk

- Splunk
  - 日志进入系统时不抽取关键字段，直接做索引，在检索时抽取关键字段
  - 灵活、索引文件小，但检索延时大
- 日志易
  - 日志进入系统时就抽取关键字段，然后做索引
  - 索引文件大，但检索延时小
  - 可在索引前配置解析规则，抽取日志里的任何关键字段

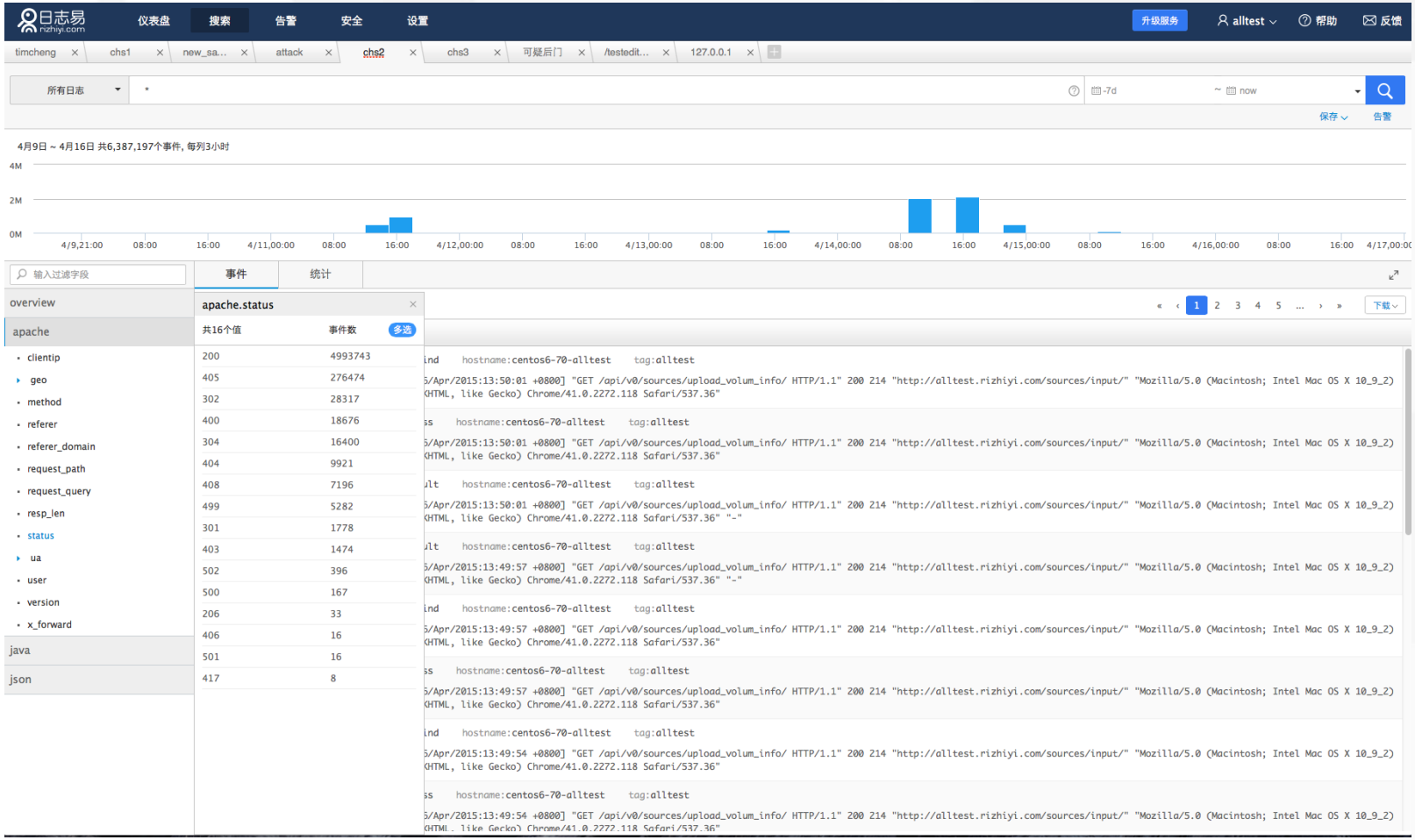
日志易介绍：总览


日志易
仪表盘
搜索
告警
安全
设置
升级服务
alltest
帮助
反馈
timcheng
chs1
new_sa...
attack
chs2
chs3
高危行为
SQL注入
所有日志
-1h
now
保存
告警
4月15日 共3,657个事件, 每列1分钟
400
200
0
4/15,17:47
17:55
18:00
18:05
18:10
18:15
18:20
18:25
18:30
18:35
18:40
18:45
输入过滤字段
事件
统计
overview
apache
clientip
geo
method
referer
referer_domain
request_path
request_query
resp_len
status
ua
version
列表
每页20
降序
升序
下载
timestamp
日志
2015-04-15 18:46:39.000
appname:alltest-default hostname:centos6-70-alltest tag:alltest
hostname: centos6-70-alltest
appname: alltest-default
tag: alltest
logtype: apache
apache:
timestamp: 1429094799000
raw_message: 192.168.1.157 - - [15/Apr/2015:18:46:39 +0800] "GET /api/v0/search/fields/?field=apache.geo.country:string&filters=tag:%22att%22&order=desc&page=1&query=*&size=100&sourcegroup=all&sourcegroupCn=%E6%89%80%E6%9C%89%E6%97%A5%E5%BF%97&time_range=-1d,now&type=fields HTTP/1.1" 200 36 "http://alltest.rizhiyi.com/search/?query=*&time_range=-1d%2Cnow&order=desc&size=20&page=1&sourcegroup=all&type=timeline&filters=tag%3A%22att%22&_t=1429094771868&title=attack&index=3" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3) AppleWebKit/600.5.17 (KHTML, like Gecko) Version/8.0.5 Safari/600.5.17" "-"
2015-04-15 18:46:39.000
appname:alltest-access hostname:centos6-70-alltest tag:alltest
192.168.1.157 - - [15/Apr/2015:18:46:39 +0800] "GET /api/v0/search/statis/fieldcategory/?filters=tag:%22att%22&method=top&order=desc&page=1&query=*&size=20&sourcegroup=all&sourcegroupCn=%E6%89%80%E6%9C%89%E6%97%A5%E5%BF%97&time_range=-1d,now&topn=5&trend=false&type=timeline HTTP/1.1" 200 36 "http://alltest.rizhiyi.com/search/?query=*&time_range=-1d%2Cnow&order=desc&size=20&page=1&sourcegroup=all&type=timeline&filters=tag%3A%22att%22&_t=1429094771868&title=attack&index=3" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3) AppleWebKit/600.5.17 (KHTML, like Gecko) Version/8.0.5 Safari/600.5.17"
2015-04-15 18:46:39.000
appname:alltest-combind hostname:centos6-70-alltest tag:alltest
192.168.1.179 - - [15/Apr/2015:18:46:39 +0800] "GET /api/v0/sources/upload_volum_info/ HTTP/1.1" 200 211 "http://alltest.rizhiyi.com/sources/input/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.118 Safari/537.36"
2015-04-15 18:46:39.000
appname:alltest-access hostname:centos6-70-alltest tag:alltest
192.168.1.157 - - [15/Apr/2015:18:46:39 +0800] "GET /api/v0/search/fields/?field=apache.geo.country:string&filters=tag:%22att%22&order=desc&page=1&query=*&size=100&sourcegroup=all&sourcegroupCn=%E6%89%80%E6%9C%89%E6%97%A5%E5%BF%97&time_range=-1d,now&type=fields HTTP/1.1" 200 36 "http://alltest.rizhiyi.com/search/?query=*&time_range=-1d%2Cnow&order=desc&size=20&page=1&sourcegroup=all&type=timeline&filters=tag%3A%22att%22&_t=1429094771868&title=attack&index=3" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3) AppleWebKit/600.5.17 (KHTML, like Gecko) Version/8.0.5 Safari/600.5.17"
2015-04-15 18:46:39.000
appname:alltest-combind hostname:centos6-70-alltest tag:alltest
192.168.1.157 - - [15/Apr/2015:18:46:39 +0800] "POST /api/v0/search/action/ HTTP/1.1" 200 25 "http://alltest.rizhiyi.com/search/?query=*&time_range=-1d%2Cnow&order=desc&size=20&page=1&sourcegroup=all&type=timeline&filters=tag%3A%22att%22&_t=1429094771868&title=attack&index=3" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10

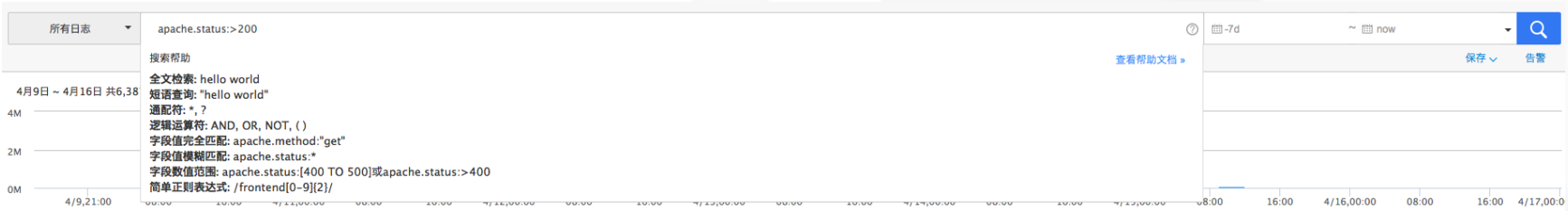
日志易介绍：日志结构化


事件
统计
列表
每页20
降序
升序
1
2
3
4
5
下载
timestamp
日志
2015-04-16 13:50:01.000
appname:alltest-combind hostname:centos6-70-alltest tag:alltest
hostname: centos6-70-alltest
appname: alltest-combind
tag: alltest
logtype: apache
apache:
clientip: 192.168.1.179
geo:
method: GET
referer: http://alltest.rizhiyi.com/sources/input/
referer_domain: alltest.rizhiyi.com
request_path: /api/v0/sources/upload_volum_info/
resp_len: 214
status: 200
ua:
browser: Chrome
browser_v: Chrome 41.0.2272
device: Other
os: Mac OS X
os_v: Mac OS X 10.9.2
version: 1.1
timestamp: 1429163401000
raw_message: 192.168.1.179 - - [16/Apr/2015:13:50:01 +0800] "GET /api/v0/sources/upload_volum_info/ HTTP/1.1" 200 214 "http://alltest.rizhiyi.com/sources/input/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.118 Safari/537.36"

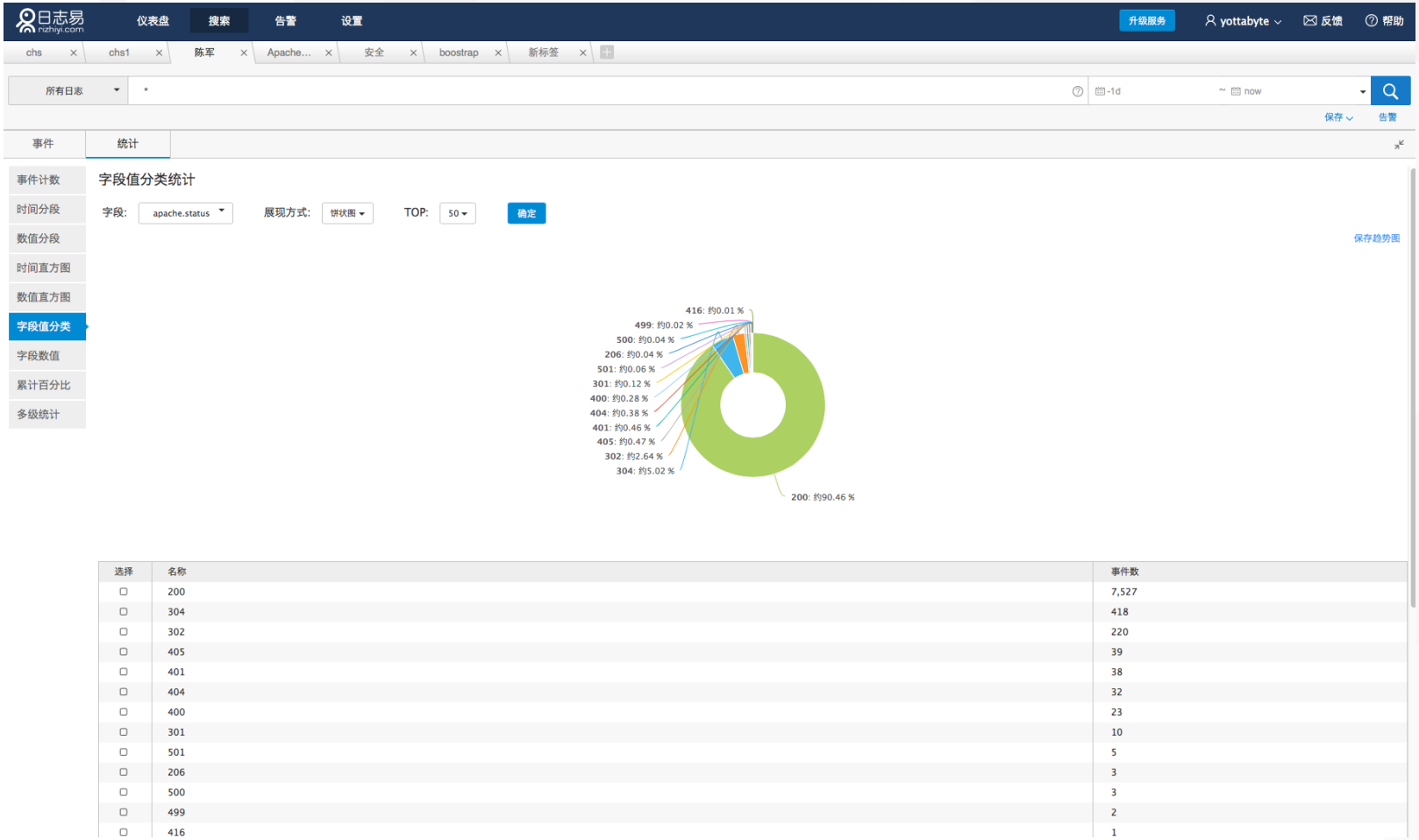
日志易介绍：字段抽取、统计



日志易介绍：搜索


所有日志
apache.status:>200
-7d
now
搜索帮助
查看帮助文档 »
保存
告警
4月9日 ~ 4月16日 共6,38
全文检索: hello world
短语查询: "hello world"
通配符: *, ?
逻辑运算符: AND, OR, NOT, ( )
字段值完全匹配: apache.method:"get"
字段值模糊匹配: apache.status:*
字段数值范围: apache.status:[400 TO 500]或apache.status:>400
简单正则表达式: /frontend[0-9]{2}/
4M
2M
0M
4/9,21:00
16:00
4/16,00:00
08:00
16:00
4/17,00:0

日志易介绍：统计

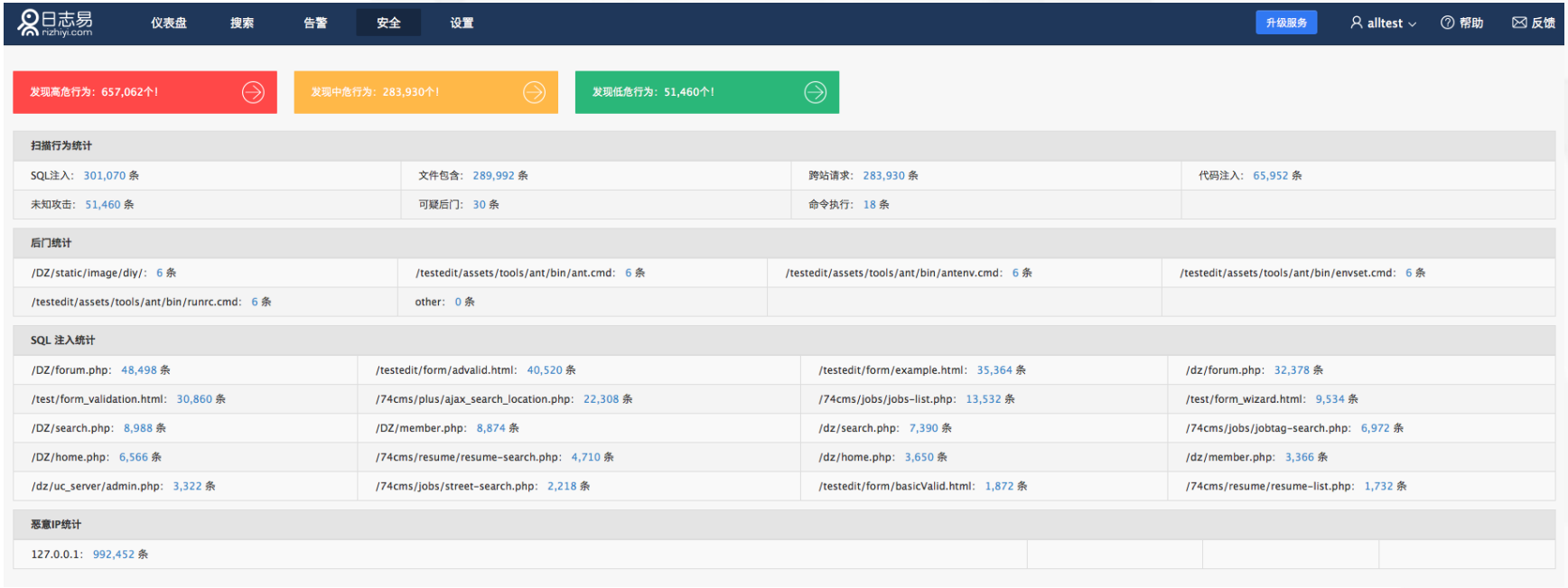


选择	名称	事件数
☐	200	7,527
☐	304	418
☐	302	220
☐	405	39
☐	401	38
☐	404	32
☐	400	23
☐	301	10
☐	501	5
☐	206	3
☐	500	3
☐	499	2
☐	416	1

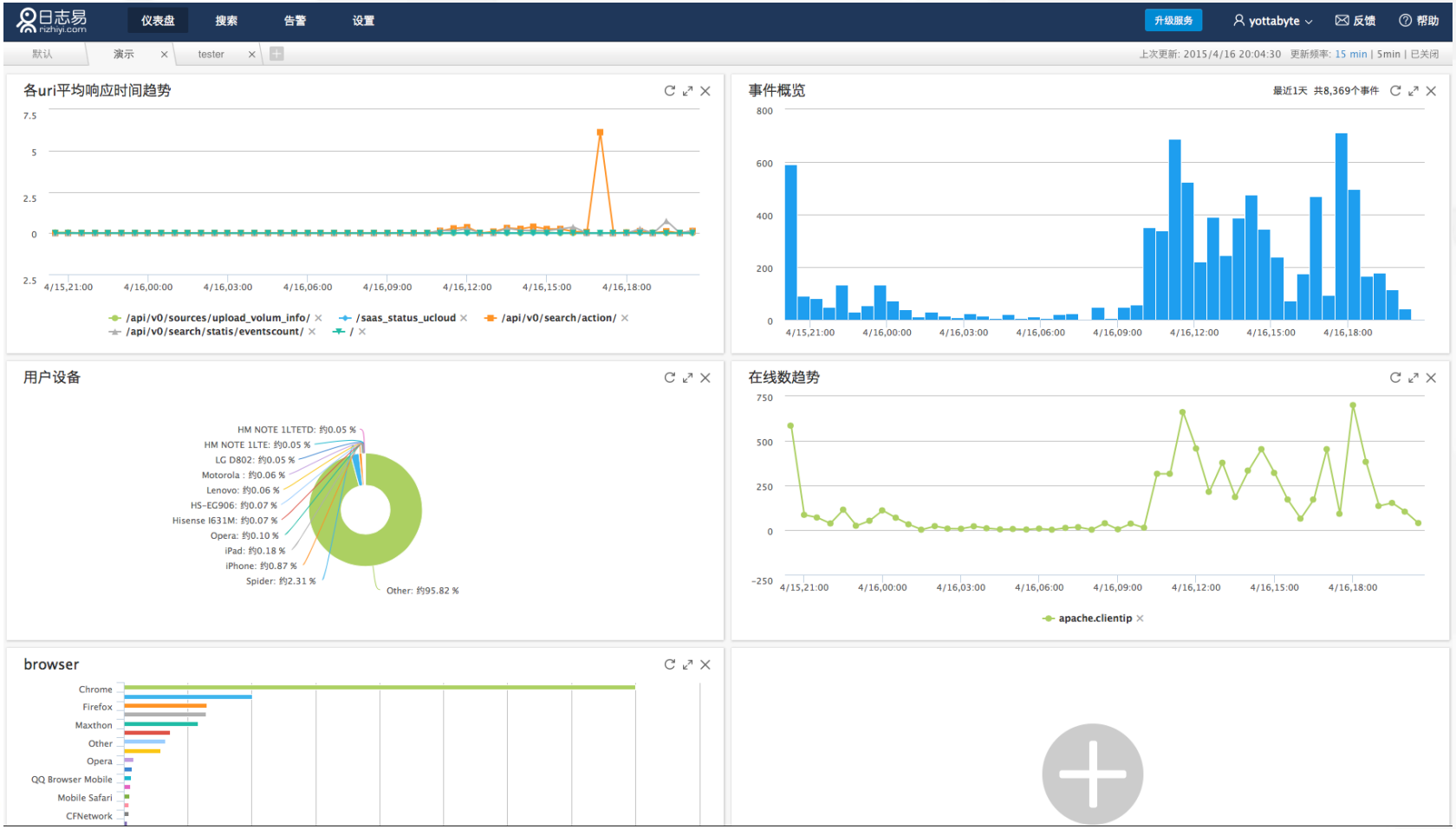
日志易介绍：告警

启用	名称	拥有者	频率	上次运行时间	24小时结果
☑	apache >= 400	yottabyte	5分钟	2015-04-16 20:01:18	
☐	alert_error	demo	1分钟	2015-03-06 19:03:54	
☑	alert_attack	yottabyte	600分钟	2015-04-16 17:13:18	
☑	apache status 错误	yottabyte	5分钟	2015-04-16 20:01:18	
☑	测试	yottabyte	5分钟	2015-04-16 20:01:19	
☑	tester01	yottabyte	5分钟	2015-04-16 20:01:19	
☑	alert_error1	yottabyte	5分钟	2015-04-16 20:01:19	

日志易介绍：安全审计


日志易 rizhiyi.com
仪表盘
搜索
告警
安全
设置
升级服务
alltest
帮助
反馈
发现高危行为：657,062个！
发现中危行为：283,930个！
发现低危行为：51,460个！
扫描行为统计
SQL注入： 301,070 条
文件包含： 289,992 条
跨站请求： 283,930 条
代码注入： 65,952 条
未知攻击： 51,460 条
可疑后门： 30 条
命令执行： 18 条
后门统计
/DZ/static/image/div/： 6 条
/testedit/assets/tools/ant/bin/ant.cmd： 6 条
/testedit/assets/tools/ant/bin/antenv.cmd： 6 条
/testedit/assets/tools/ant/bin/envset.cmd： 6 条
/testedit/assets/tools/ant/bin/runrc.cmd： 6 条
other： 0 条
SQL 注入统计
/DZ/forum.php： 48,498 条
/testedit/form/advalid.html： 40,520 条
/testedit/form/example.html： 35,364 条
/dz/forum.php： 32,378 条
/test/form_validation.html： 30,860 条
/74cms/plus/ajax_search_location.php： 22,308 条
/74cms/jobs/jobs-list.php： 13,532 条
/test/form_wizard.html： 9,534 条
/DZ/search.php： 8,988 条
/DZ/member.php： 8,874 条
/dz/search.php： 7,390 条
/74cms/jobs/jobtag-search.php： 6,972 条
/DZ/home.php： 6,566 条
/74cms/resume/resume-search.php： 4,710 条
/dz/home.php： 3,650 条
/dz/member.php： 3,366 条
/dz/uc_server/admin.php： 3,322 条
/74cms/jobs/street-search.php： 2,218 条
/testedit/form/basicValid.html： 1,872 条
/74cms/resume/resume-list.php： 1,732 条
恶意IP统计
127.0.0.1： 992,452 条

日志易介绍：仪表盘


日志易
仪表盘
搜索
告警
设置
升级服务
yottabyte
反馈
帮助
默认
演示
tester
上次更新: 2015/4/16 20:04:30 更新频率: 15 min | 5min | 已关闭
各uri平均响应时间趋势
/api/v0/sources/upload_volum_info/
/saas_status_ucloud
/api/v0/search/action/
/api/v0/search/statis/eventscount/
/
事件概览
最近1天 共8,369个事件
用户设备
HM NOTE 1LTETD: 约0.05 %
HM NOTE 1LTE: 约0.05 %
LG D802: 约0.05 %
Motorola : 约0.06 %
Lenovo: 约0.06 %
HS-EG906: 约0.07 %
Hisense I631M: 约0.07 %
Opera: 约0.10 %
iPad: 约0.18 %
iPhone: 约0.87 %
Spider: 约2.31 %
Other: 约95.82 %
在线数趋势
apache.clientip
browser
Chrome
Firefox
Maxthon
Other
Opera
QQ Browser Mobile
Mobile Safari
CFNetwork

日志易，日志分析更容易

rizhiyi.com