



# 2016 全球运维大会

深圳站

会议时间：3月25日-3月26日

会议地点：深圳·南山区 圣淘沙酒店(翡翠店)

主办单位： 开放运维联盟  
OOPSA Open OPS Alliance  高效运维社区  
Great OPS Community

指导单位： 数据中心联盟  
Data Center Alliance

协办单位：中国新一代IT产业推进联盟





# 全球运维大会 2016

深圳站

## IT运维分析与海量日志搜索

陈军，日志易CEO



# 提纲

- ▶ IT 运维分析 ( IT Operation Analytics )
- ▶ 日志的应用场景
- ▶ 过去及现在的做法
- ▶ 日志搜索引擎
- ▶ 日志易产品介绍



# IT 运维分析

- ✦ 从 IT Operation Management (ITOM) 到 IT Operation Analytics (ITOA)
- ✦ 大数据技术应用于IT运维，通过数据分析提升IT运维效率
  - 可用性监控
  - 应用性能监控
  - 故障根源分析
  - 安全审计
- ✦ Gartner估计，到2017年15%的大企业会积极使用ITOA；而在2014年这一数字只有5%

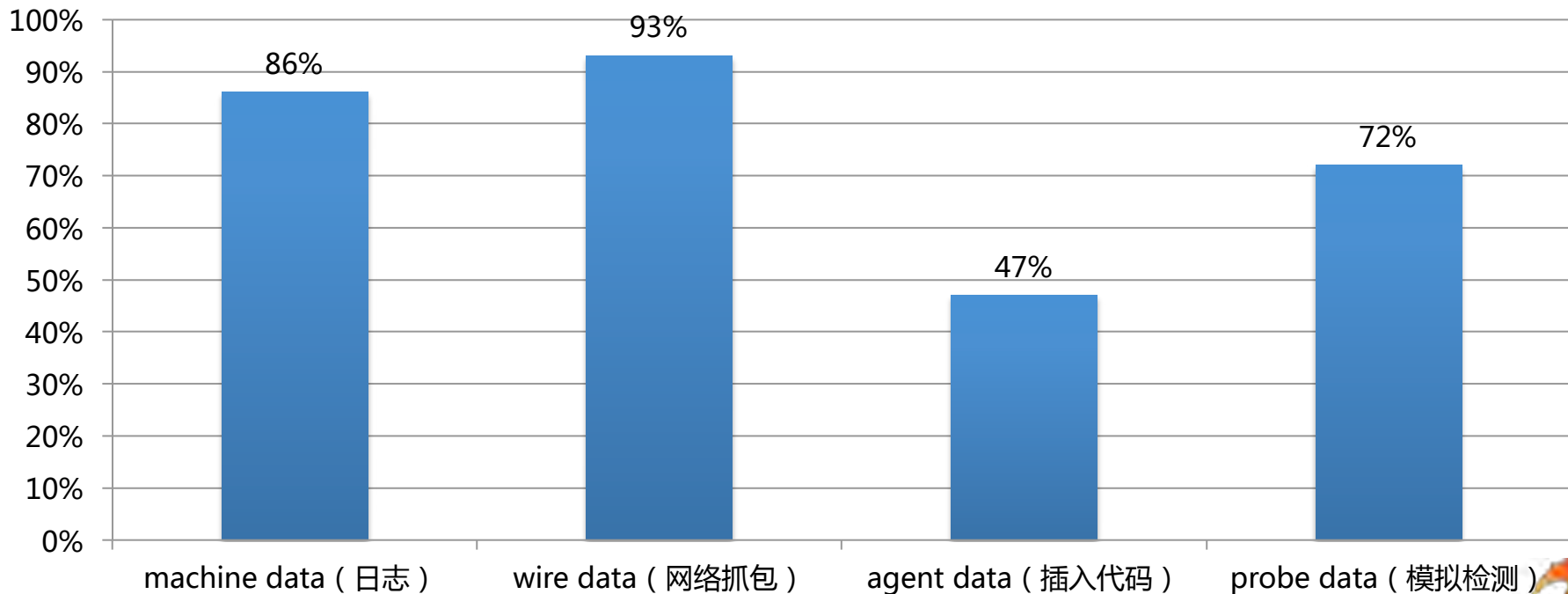


# ITOA 的四种数据来源

- ✦ 机器数据 ( Machine Data )
  - 日志
- ✦ 通信数据 ( Wire Data )
  - 网络抓包, 流量分析
- ✦ 代理数据 ( Agent Data )
  - 在 .NET/Java 字节码里插入代码, 统计函数调用、堆栈使用
- ✦ 探针数据 ( Probe Data )
  - 在各地模拟ICMP ping、HTTP GET请求, 对系统进行检测



# ITOA 四种数据来源使用占比



# ITOA 四种数据来源的比较

- ✦ 机器数据（日志）
  - 日志无所不在
  - 但不同应用输出的日志内容的完整性、可用性不同
- ✦ 通信数据（网络抓包）
  - 网络流量信息全面
  - 但一些事件未必触发网络流量
- ✦ 代理数据（嵌入代码）
  - 代码级精细监控
  - 但侵入性，会带来安全、稳定、性能问题
- ✦ 探针数据（模拟用户请求）
  - 端到端监控
  - 但不是真实用户度量（Real User Measurement）



# 日志：时间序列机器数据

- ✦ 带时间戳的机器数据
- ✦ IT 系统信息
  - 服务器
  - 网络设备
  - 操作系统
  - 应用软件
- ✦ 用户信息
  - 用户行为
- ✦ 业务信息
- ✦ 日志反映的是事实数据
  - “The Log: What every software engineer should know about real-time data's unifying abstraction” , Jay Kreps, LinkedIn engineer
  - 深度解析LinkedIn大数据平台 ( <http://www.csdn.net/article/2014-07-23/2820811/1> )





# 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



# 日志的应用场景

## ✦ 运维监控

- 可用性监控
- 应用性能监控 (APM)

## ✦ 安全审计

- 安全信息事件管理 (SIEM)
- 合规审计
- 发现高级持续威胁 (APT)

## ✦ 用户及业务统计分析



# 过去

- ✦ 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



# 近年

## ✦ Hadoop

- 批处理，不够及时
- 查询慢
- 数据离线挖掘，无法做 OLAP (On Line Analytic Processing)

## ✦ Storm/Spark

✦ Hadoop/Storm/Spark都只是一个开发框架，不是拿来即用的产品

## ✦ NoSQL

- 不支持全文检索



# 现在

- ✦ 对日志实时搜索、分析
  - 日志实时搜索分析引擎
- ✦ 快
  - 日志从产生到搜索分析出结果只有几秒的延时
- ✦ 大
  - 每天处理 TB 级的日志量
- ✦ 灵活
  - Google for IT , 可搜索、分析任何日志
- ✦ Fast Big Data



# 日志管理系统的进化



# 日志易亮点

- ✦ 可编程的日志实时搜索分析平台
- ✦ 搜索处理语言（ Search Processing Language, SPL ）
  - SPL命令用管道符（ “|” ）串接成脚本程序
  - 在搜索框里写 SPL 脚本，完成复杂的查询、分析
- ✦ 可接入各种来源的数据
  - 日志文件
  - 数据库
  - 恒生电子交易系统二进制日志
- ✦ 企业部署版
- ✦ SaaS 版
  - 每天500MB日志处理免费



# 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
  - 把日志从非结构化数据转换成结构化数据
- ✦ 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
  - 乐视：100万 EPS (Event Per Second)，20TB/天





# 客户案例：某大型综合金融机构

## ✦ 使用日志易之前

- 逐台登陆服务器，无法集中查看日志，无法对海量数据进行挖掘、用户行为分析
- 日志查询方式比较原始，只能 less、grep 和 awk 等常见的 Linux 指令，无法多维度查询（时间段、关键字、字段值）
- 无法进行日志的业务逻辑分析和告警

## ✦ 使用日志易之后，接入100多个应用的日志，8TB/天

- 省去登陆服务器的操作，快速，降低人为登陆服务器误操作引发生产故障
- 查询条件多维度，提升定位异常原因的效率
- 可以对日志数据进行数据挖掘、用户行为分析并产生相应的报表，同时还可以针对应用系统健康指数提前告警，而不是事后补漏



# 客户案例：中移动某省分公司

## ✦ 使用场景和解决的问题

- 分析营业厅业务办理Web请求日志
- 聚合出每个营业员每项业务的详细操作步骤，对每个步骤操作时长进行告警、统计分析

## ✦ Search Processing Language 范例

```
json.url:"/charge/business.action?BMEBusiness=charge.charge&_cntRecTimeFlag=true" | transaction  
apache.dimensions.cookie_CURRENT_MENUID startswith=eval(json.action:"查询" &&  
timestamp<30m) endswith=json.action:"提交"
```

1.先通过url过滤出所有缴费业务日志

5.将“提交”动作作为步骤结束

2.通过menuid进行分组聚合

3.将“查询”动作作为步骤起点

4.默认30分钟内营业员处理完一笔完整业务



# 客户案例：中移动某省分公司

日志易 仪表盘 搜索 首页 应用 设置 admin 帮助

各地址... x 业务起... x

前页 后页 [json dimensions cookie\_CURRENT\_MENUID="BIAR\_Change\_WFB"]transaction [json dimensions cookie\_Login\_Go 2015/10/17 12:30:00.0 2015/10/17 13:30:00.0

过滤字段: [log\_compuware] x

可执行

列表 表格 树形 详细 每页20

timestamp	json dimensions c	日志
	appName: userAction log: compuware logtype: json	
	actionName: clickOn * * * _Cad_ keypress<CTRL>RN> on factPay clickOn * * * application: www.zzidboss.com www.zzidboss.com www.zzidboss.com www.zzidboss.com clientError: 0 0 0 0 cpuTime: 103.71712618083954 30.007210072221236 13.308280015021806 23.088551188600123	
	dimensions IP: 134.45.209.210 134.45.209.210 134.45.209.210 134.45.209.210 cookie_CURRENT_MENUID: BIAR_Change_WFB BIAR_Change_WFB BIAR_Change_WFB BIAR_Change_WFB cookie_Login_Cookie: n5230005 n5230005 n5230005 n5230005	
	duration: 23.088551188600123 30.007210072221236 13.308280015021806 103.71712618083954 execTime: 3051.6505530178517 1792.0781377058281 103.0020230151885 11007.01075158407 failed: false false false false	
	measures Network_Contribution: 36.344183543757026 23.760257691880486 3.2650923766152022 16.013561899120015 Server_Contribution: 858.0027942657471 102.04208374023438 15.048831939697266 52.33828163146973 name: 用户操作时间n523 用户操作时间n523 用户操作时间n523 用户操作时间n523 sumPathId: PT=266159064;PA=-1281484067;PS=-1092515210 PT=266159064;PA=-1281484067;PS=-1092515210	

一笔缴费业务营业员所有操作步骤一目了然

每个步骤所需要的执行时间按步骤顺序排列

网络处理时间，服务器处理时间按步骤顺序排列

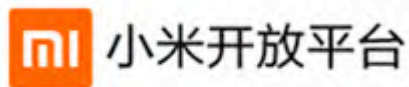
# 客户案例：国家电网

## ✦ 安全信息与事件管理

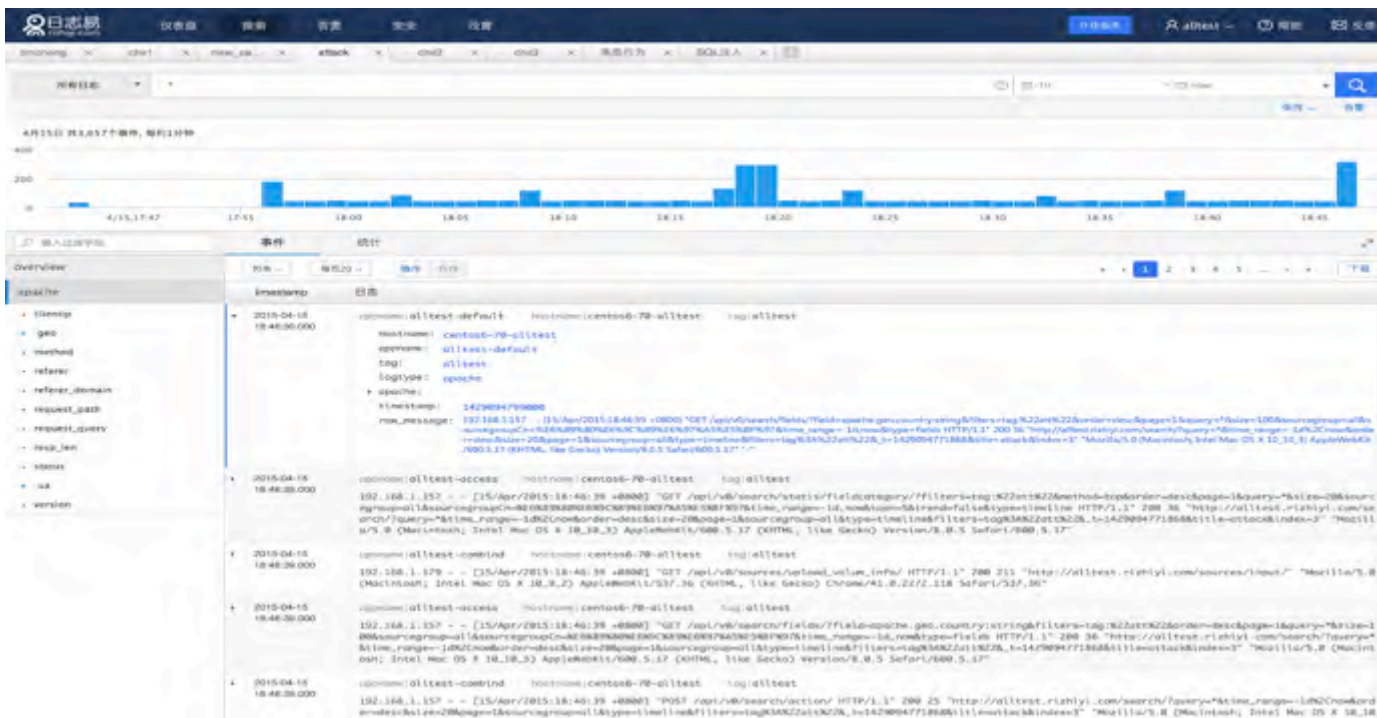
- 终端信息安全事件日志的调查、分析、取证
- 在各省分公司信息安全事件现场使用
- 快速排查事件日志保留的证据，为事件取证提供支持



# 客户



# 日志易介绍：总览

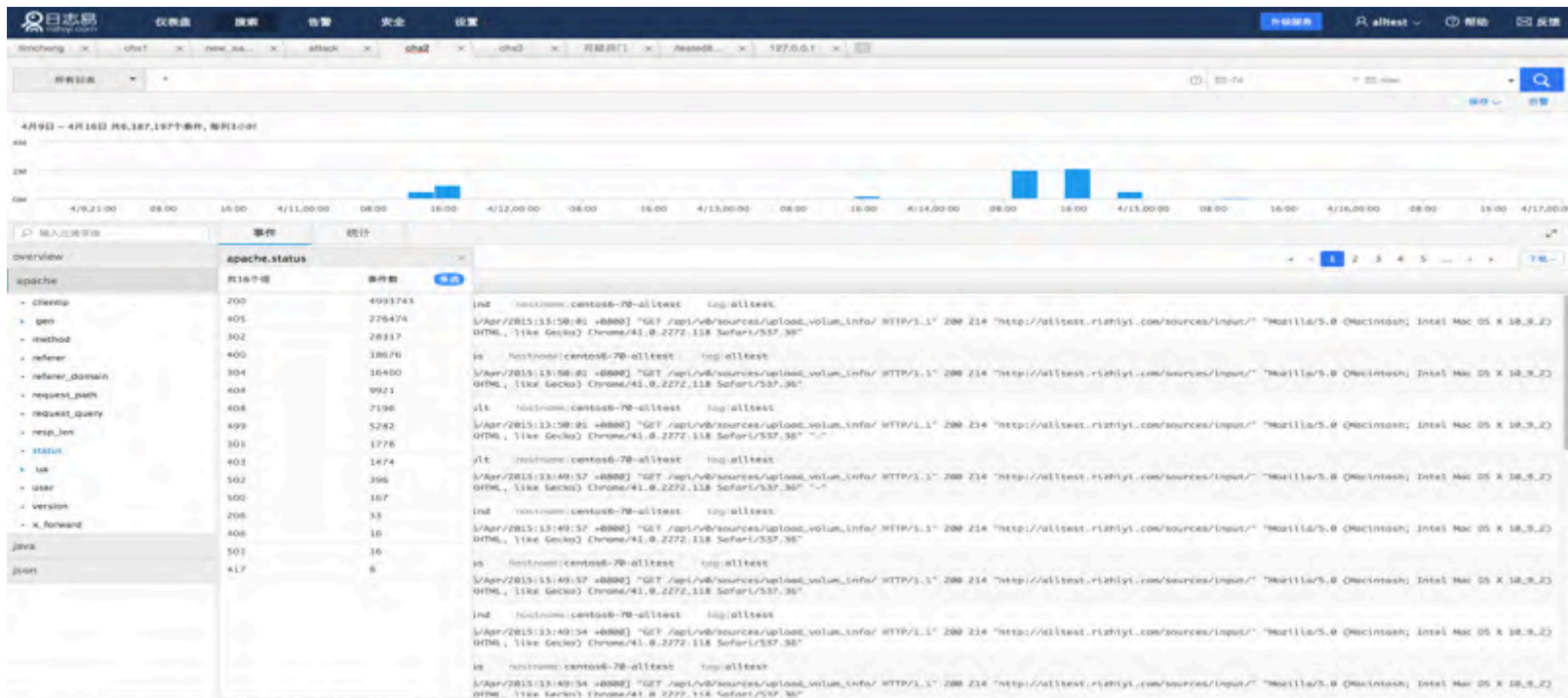


# 日志易介绍：日志结构化

事件	统计
列表	每页20 排序 升序
1 2 3 4 5 ... 下载	
Timestamp	日志
2015-04-16 13:50:01.000	<pre>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"</pre>



# 日志易介绍：字段抽取、统计

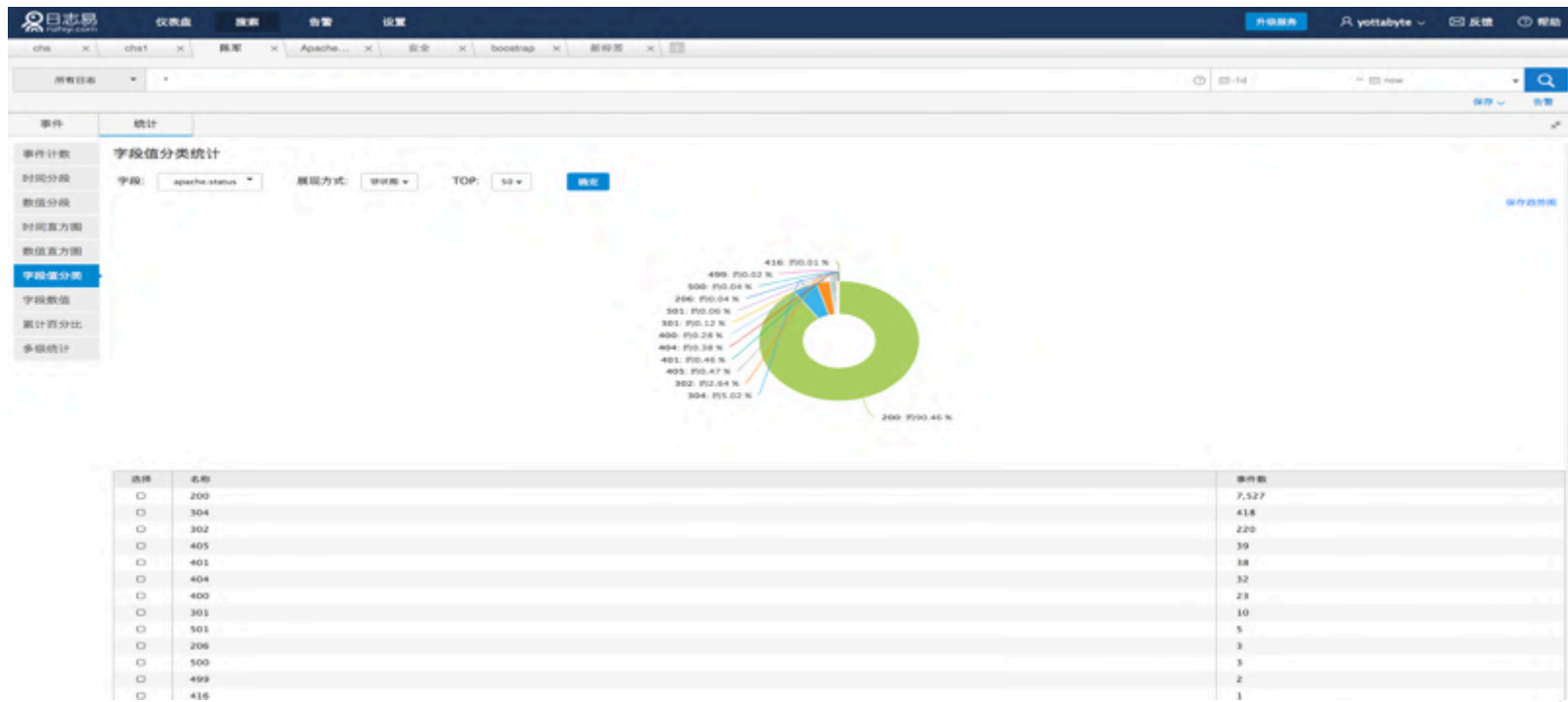




# 日志易介绍：搜索



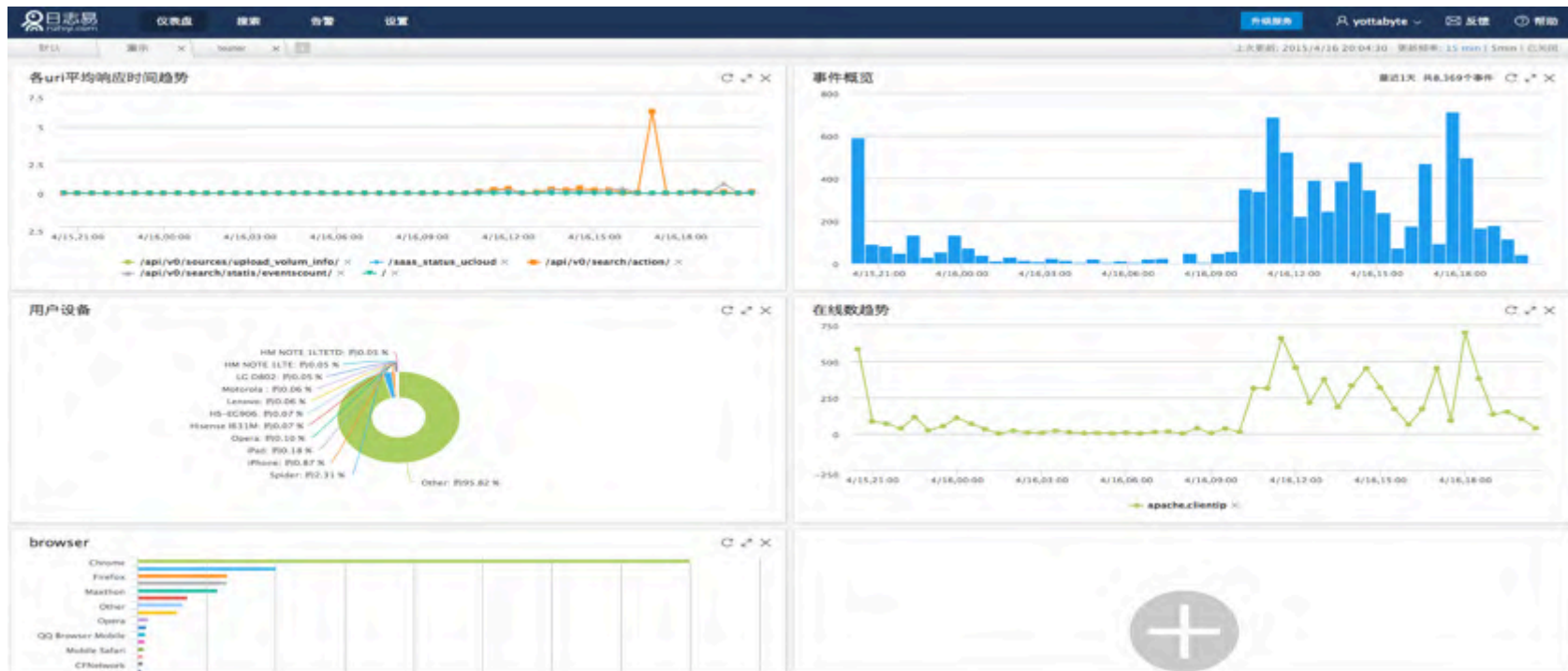
# 日志易介绍：统计



# 日志易介绍：告警



# 日志易介绍：仪表盘



# 公司情况

## ✦ 融资

- 2014年3月，获得真格基金等天使投资人1400万元天使投资
- 2015年12月，获得红杉资本6000万元A轮投资

## ✦ 团队

- 来自 BAT、360的核心研发团队
- 来自著名外企的核心销售团队





日志易，日志分析更容易！

[rizhiyi.com](http://rizhiyi.com)

