日志易
rizhiyi.com

IT运维分析与海量日志搜索

日志易CEO 陈军

日志易 rizhiyi.com

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

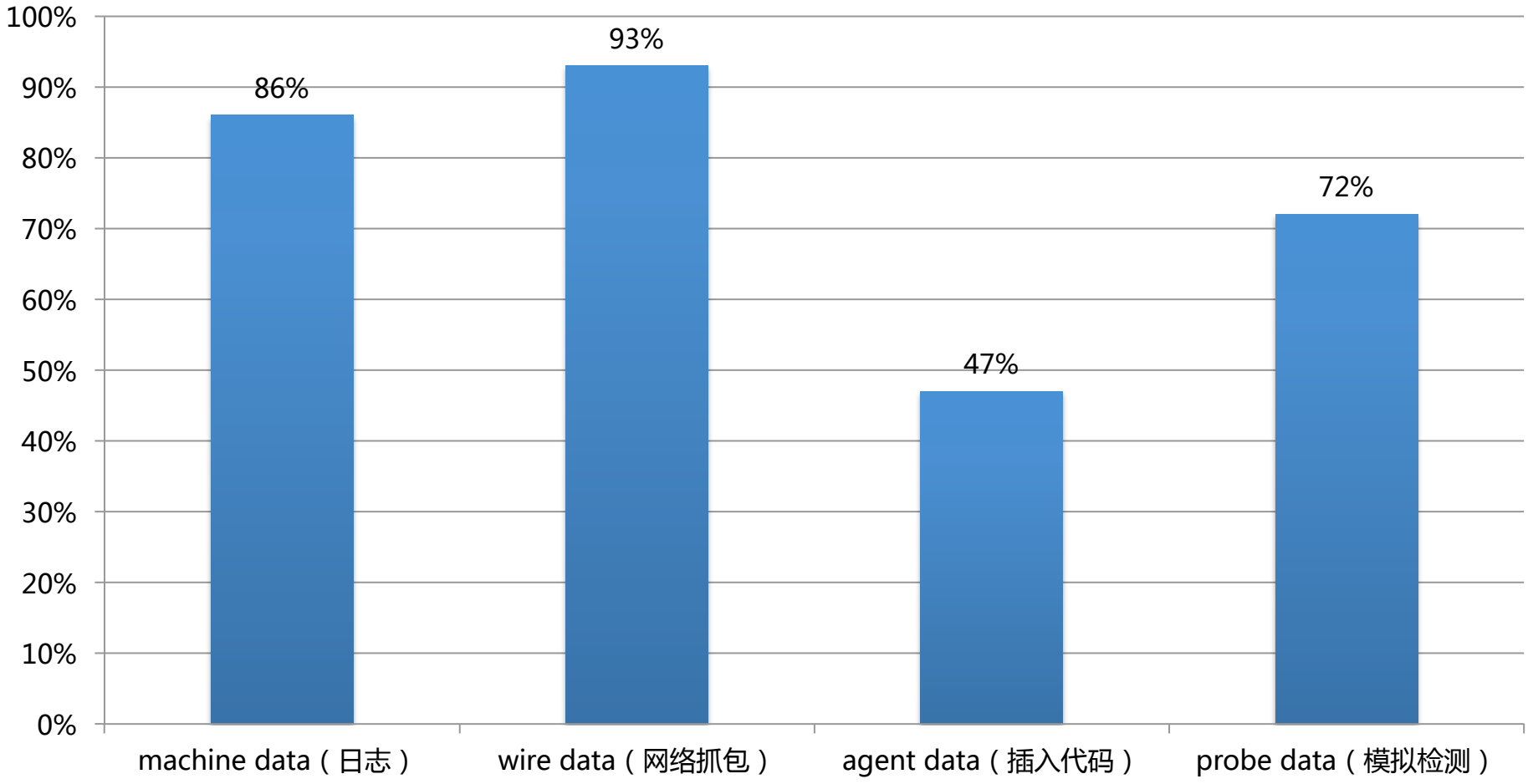
- 从 IT Operation Management (ITOM) 到 IT Operation Analytics (ITOA)
- 大数据技术应用于IT运维，通过数据分析提升IT运维效率
  - 可用性监控
  - 应用性能监控
  - 故障根源分析
  - 安全审计
- Gartner估计，到2017年15%的大企业会积极使用ITOA；而在2014年这一数字只有5%

日志易 rizhiyi.com

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，我们重要的数据资产

行为日志　　网络日志　　交易日志　　应用及系统日志

IT系统（服务器、网络设备）每天都产生大量的日志，包含了各种设备、系统、应用、用户信息

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

日志易 rizhiyi.com

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

日志易 rizhiyi.com

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

日志易 rizhiyi.com

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

日志易 rizhiyi.com

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

日志易 rizhiyi.com

日志管理系统的进化

- 固定的schema无法适应任意日志格式
- 无法处理大数据量

- 需要开发成本
- 批处理，实时性差
- 不支持全文检索

- 实时
- 灵活
- 全文检索

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

Schema on Write vs. Schema on Read

- Schema on Write
  - 索引时（入库前）抽取字段，对日志做结构化
  - 检索速度快
  - 但不够灵活，必须预先知道日志格式
- Schema on Read
  - 检索时（入库后）抽取字段，对日志结构化
  - 灵活，检索时根据需要抽取字段
  - 但检索速度受影响
- 日志易同时支持 Schema on Write 和 Schema on Read
  - 日志易实现机制
  - 由用户选择需要的策略

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
  - 索引性能：100万 EPS (Event Per Second)，20TB/天
  - 检索性能：60秒内检索1000亿条日志

日志易分析事件优势

完备的全量日志管理

日志分析的关键在于其完备性。日志易能够完整保存长周期、大容量的日志数据，为后期的分析提供了基础

细粒度的数据分析

日志的格式、内容五花八门，对其分析的方式方法更是如此。日志易提供了灵活、高效的数据分析语句，能够帮助用户从容的进行细粒度的数据分析

可视化统计

分析人员通过几下鼠标点击，即可快速完成诸如计数、时间段、数值分布、百分比、多级汇总、地理分布等统计操作，并通过最适合的图表进行呈现

秒级回馈

分析人员的任何一个想法、一个线索、一个疑点，都可以在几十甚至几秒的时间内得到验证，极大的提高了数据分析的效率

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160多个应用的日志，10TB/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

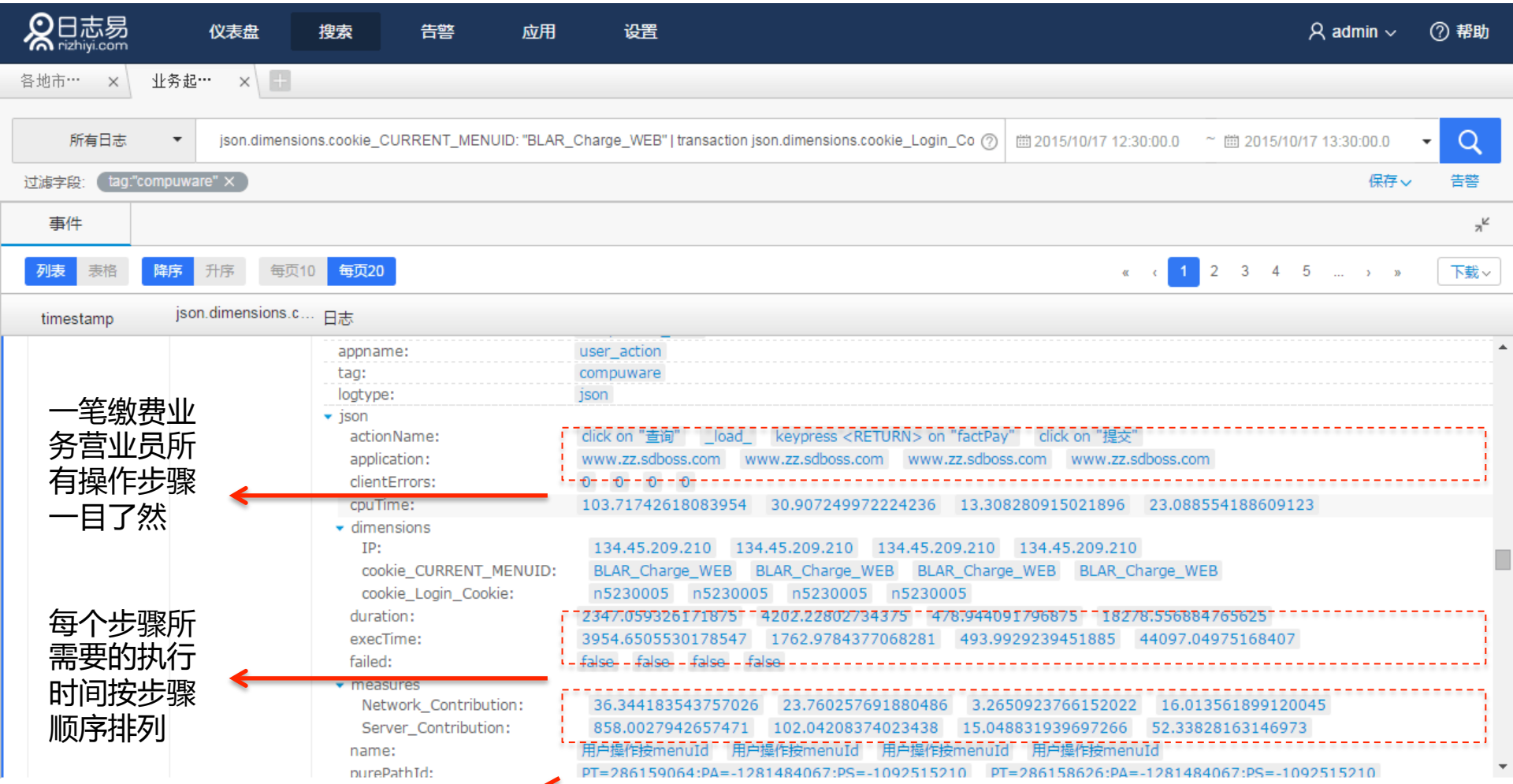
4.默认30分钟内营业员处理完一笔完整业务

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

客户案例：中移动某省分公司



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

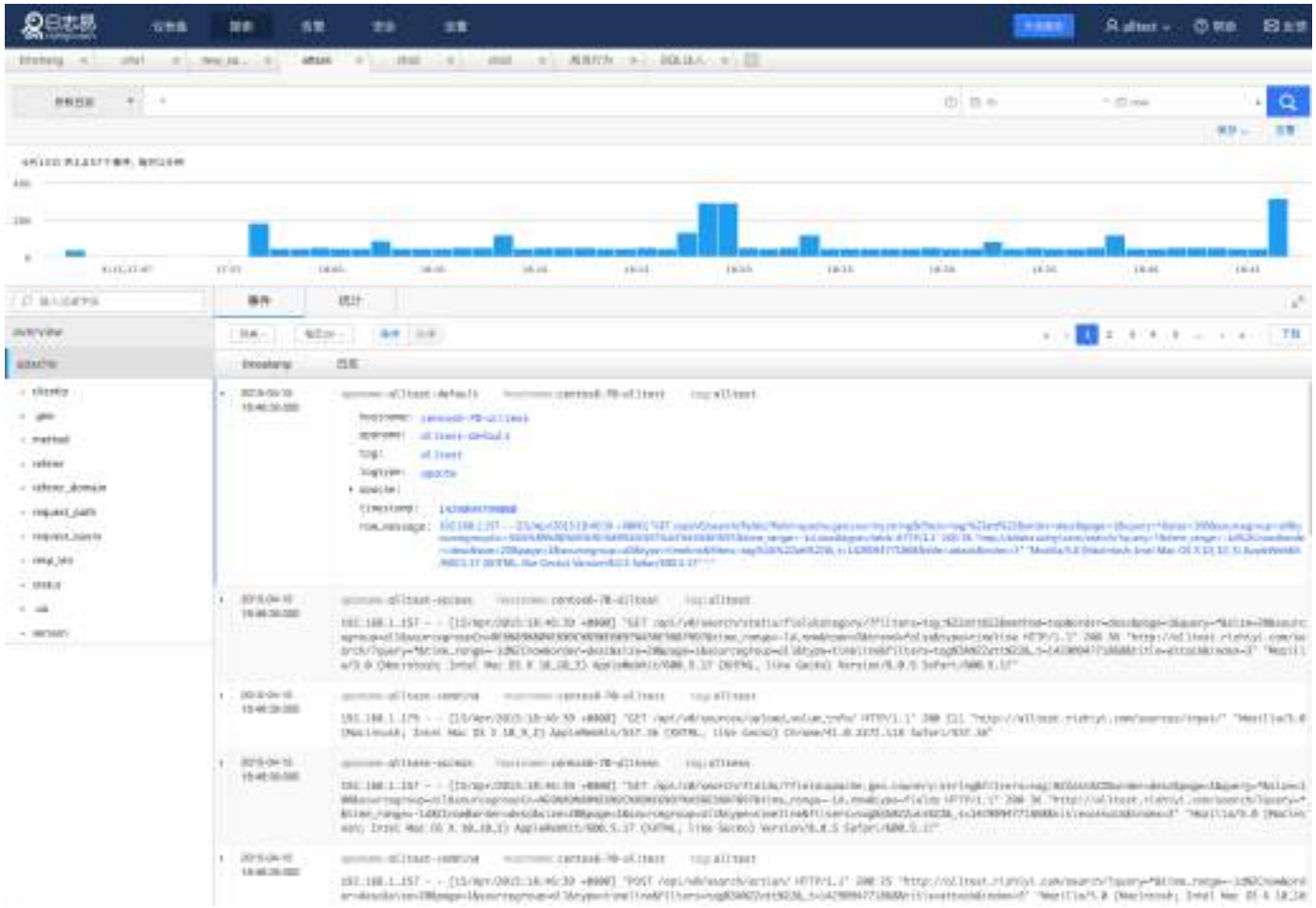
- 安全信息与事件管理
  - 终端信息安全事件日志的调查、分析、取证
  - 在各省分公司信息安全事件现场使用
  - 快速排查事件日志保留的证据，为事件取证提供支持

日志易 rizhiyi.com

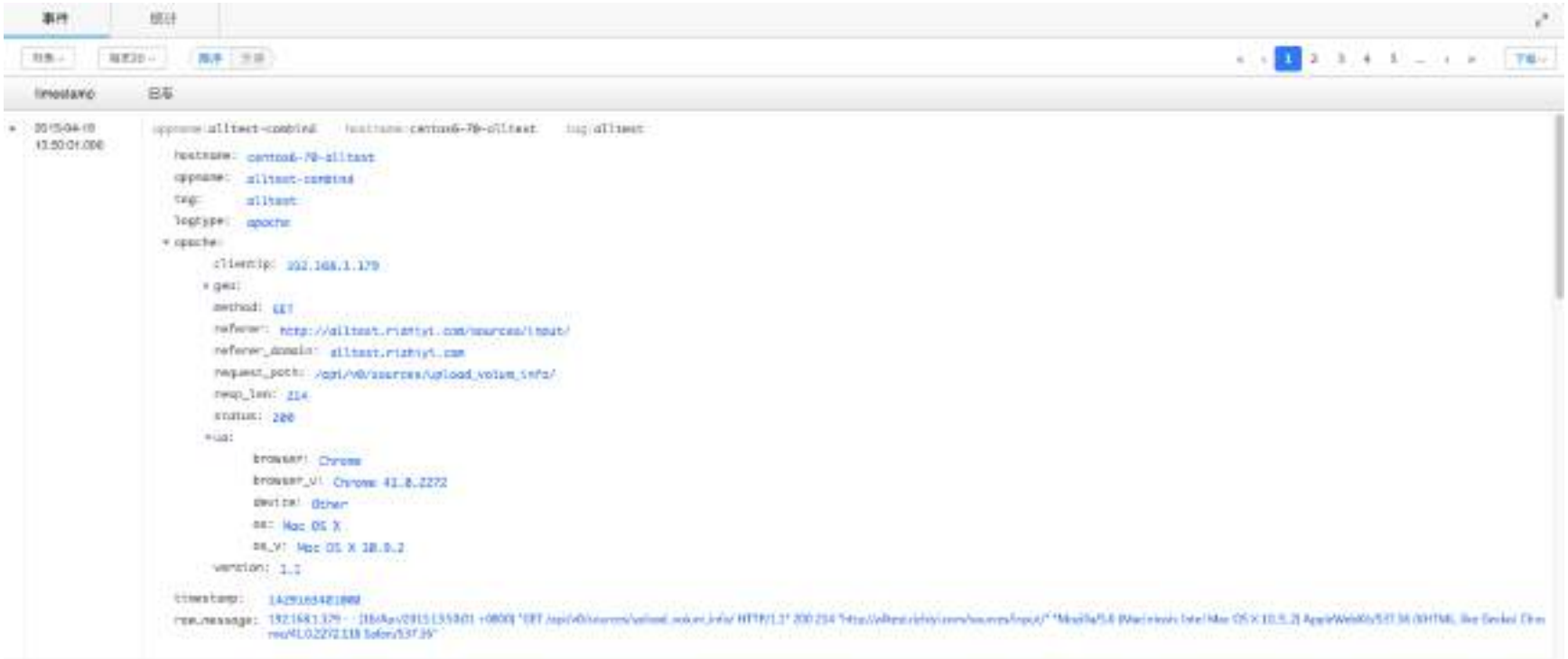
客户

日志易
rizhiyi.com

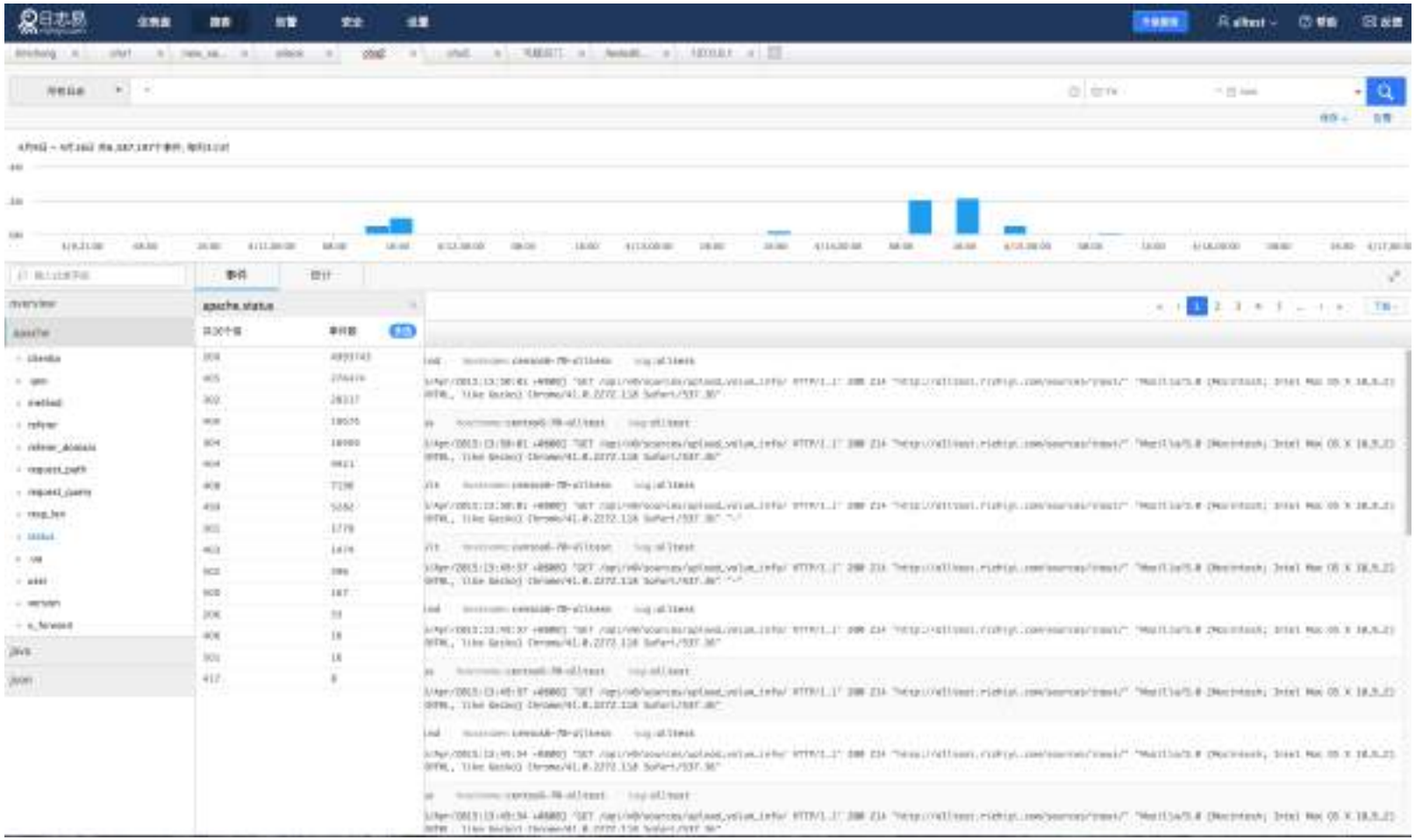
日志易介绍：总览



日志易介绍：日志结构化

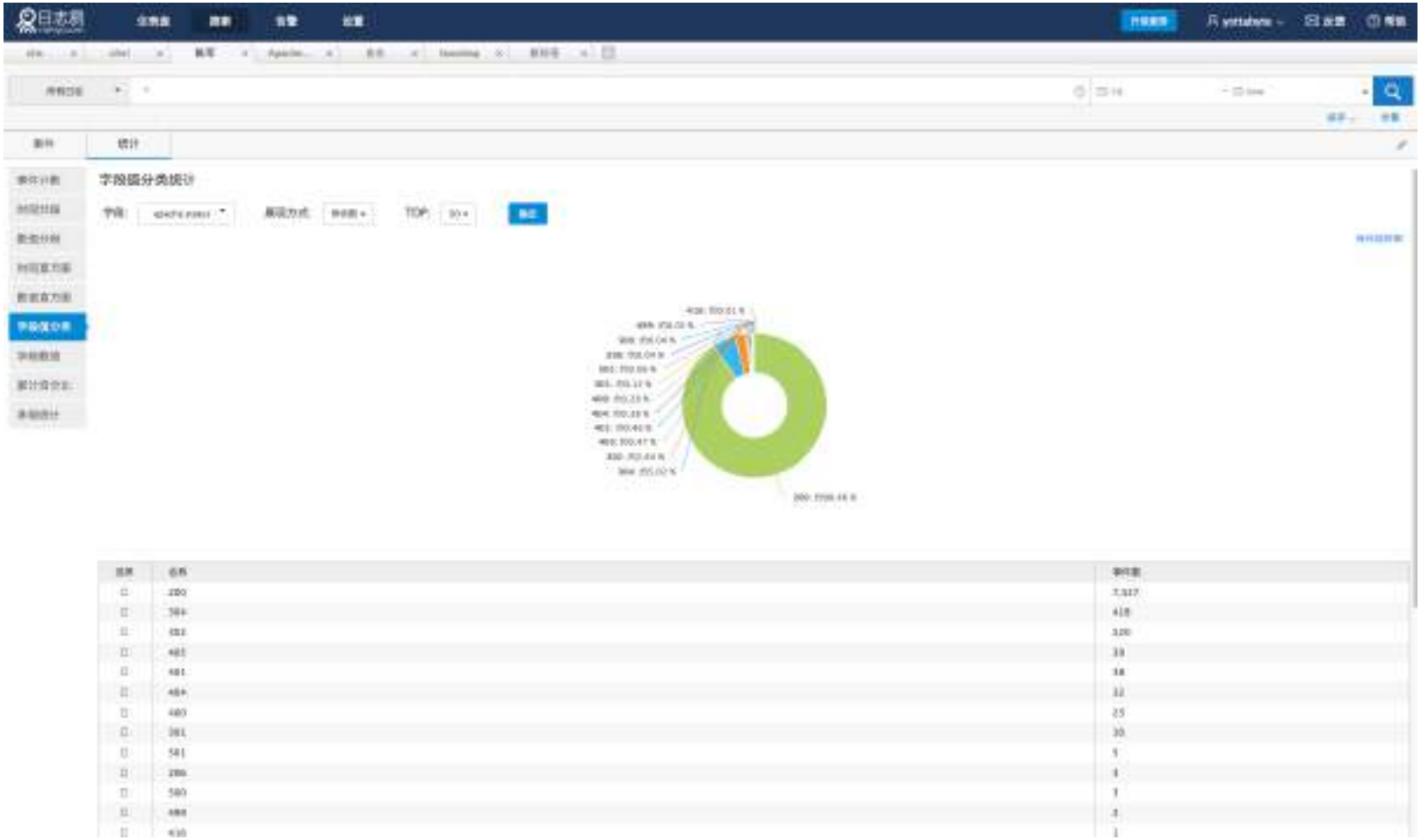


日志易介绍：字段抽取、统计



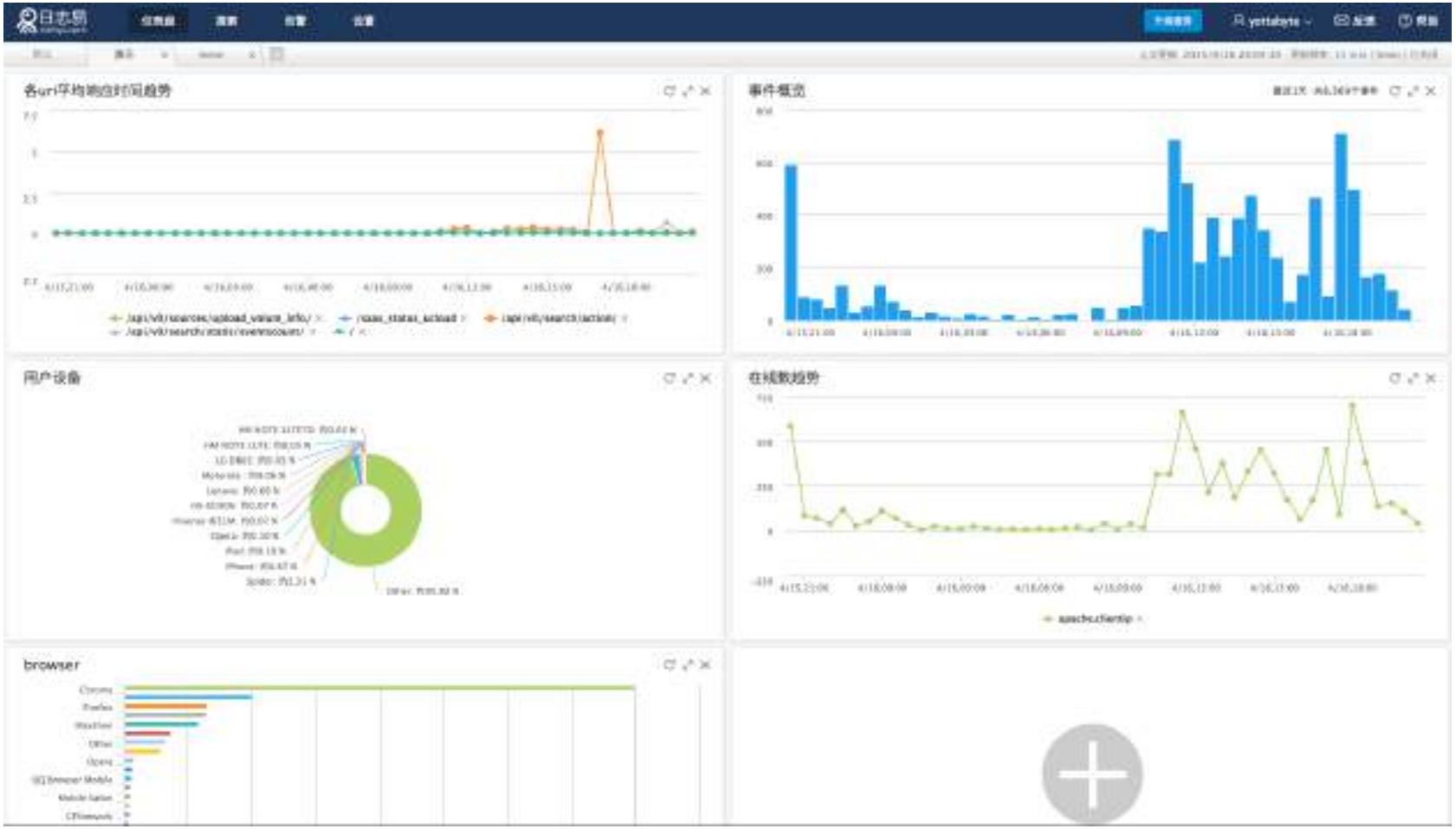
日志易介绍：搜索

日志易介绍：统计

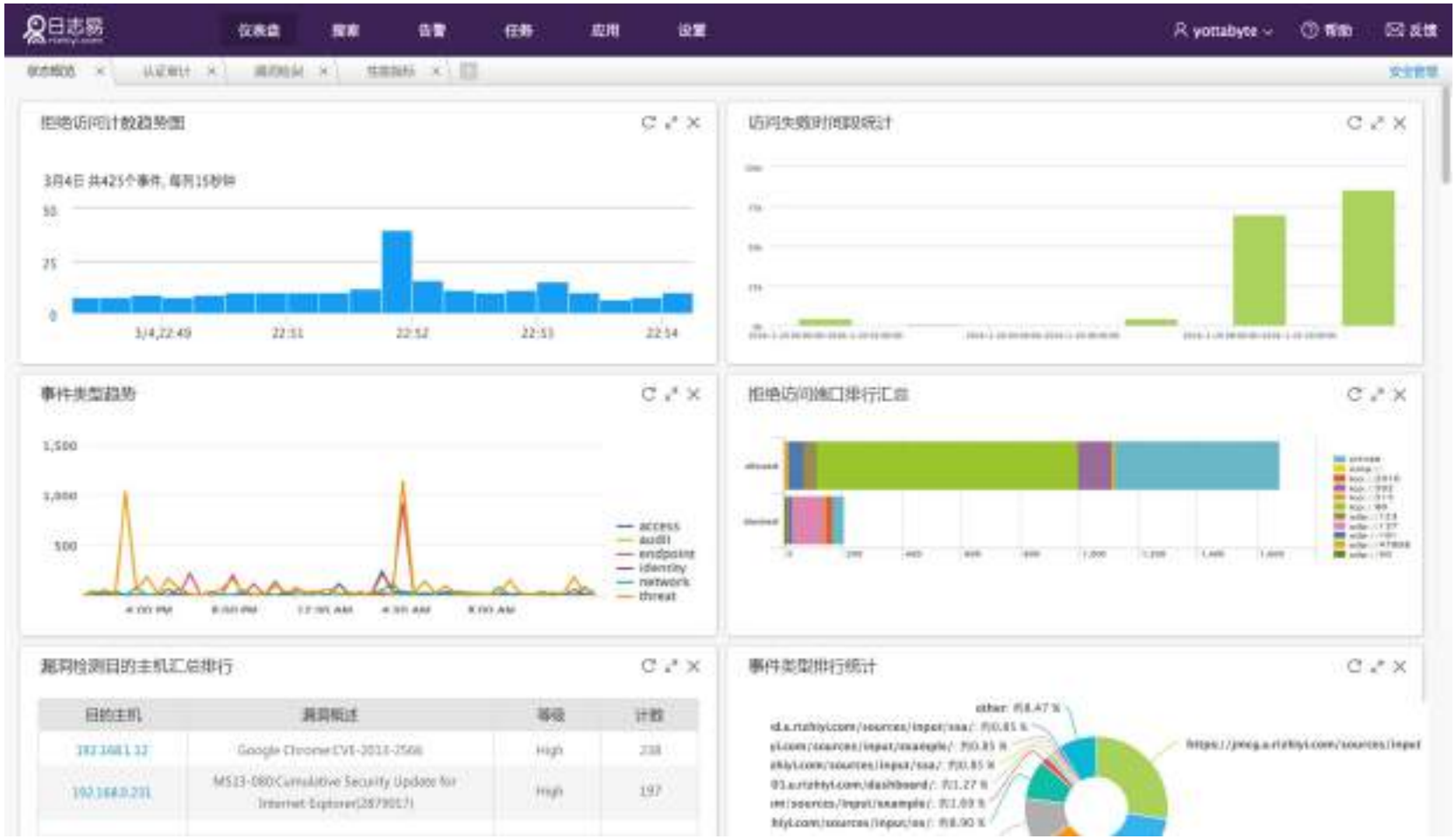


日志易介绍：告警

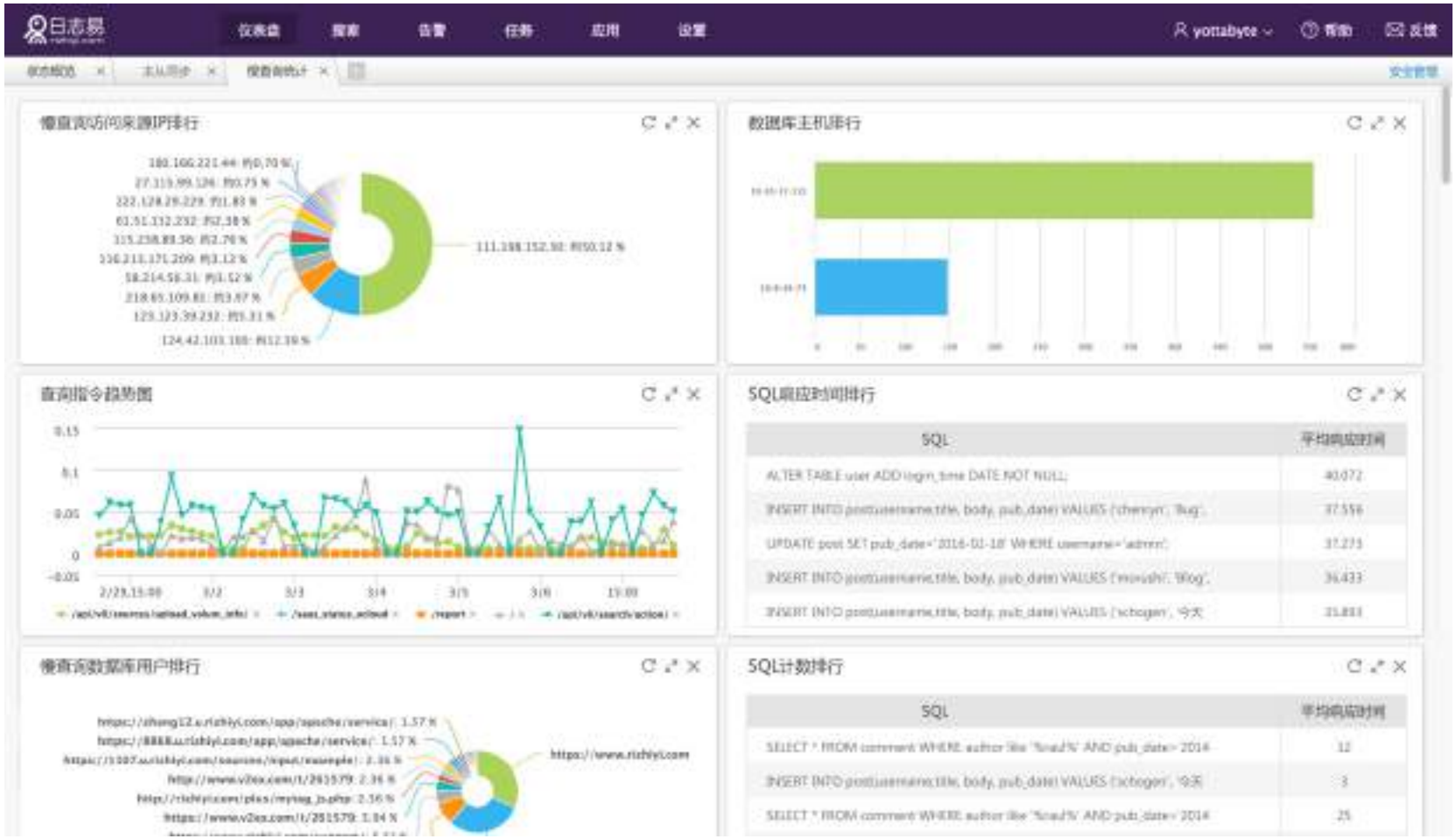
日志易介绍：仪表盘



日志易-网络安全部门仪表盘



日志易-应用监控仪表盘



日志易-网络设备部门仪表盘

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

微信公众号