日志易产品白皮书

北京优特捷信息技术有限公司

Version 2.2, 2018-08-01

序

在21世纪的今天，随着互联网和大数据时代的到来，中国的企业正面临着前所未有的挑战，互联网、金融、电信、能源及其他行业的企业都明白，公司需要专注于自己的业务。日志管理分析是基础工具，拿来就用是最省时间和资源的作法。对于已经自主开发了内部日志管理系统的公司，随着业务发展日志管理会越来越复杂，自主开发、维护日志管理分析工具的成本将会超过直接使用日志易的成本。

一些公司的日志分散在各台服务器上，每次查找日志都要登陆到各台服务器，效率低下。使用日志易统一管理日志，在一个界面上就可以查看所有日志，大大提高运维效率。

黑客在入侵服务器或网络设备时，往往会删掉日志，抹除作案证据。使用日志易统一上传、管理日志，可及时发现入侵行为，监控告警，也可以长期保存日志，方便事后安全审计。

一些公司的日志由各业务部门分别处理，导致了日志数据及分析结果的碎片化。日志是一家公司运营情况的真实数据，不同业务部门的日志往往互相关联。在公司层面统一处理、分析日志，可以把不同来源的日志对照关联分析，去除噪音，反应真实情况。

建立统一日志处理平台对一家公司的意义重大，日志易以一种崭新的“以日志搜索引擎”为核心的日志实时管理系统应运而生，会逐步成为日志分析热门领域。

1. 关于日志易

1.1. 日志管理的价值在哪里?

从传统的syslog，到正在不断兴起的云计算和大数据，对于负责运维工作的技术人员来说，日志是一个熟悉的名词——机房中的各种系统、服务器和网络设备都在不断地产生日志。每天的安全入侵和渗透攻击都希望能实时发现和防护，但我们是否能在系统出现问题或问题隐患的时候，是否及时能从日志分析中看出端倪？是否能快速定位故障及时恢复业务？健全的日志记录和分析系统是系统正常运营优化及安全事故响应的基础。而“日志管理”在企业价值提升的过程中无疑扮演着一个极为关键的角色。

1.2. 为什么选择日志易?

- **有些用户提到已经使用grep/sed/awk等Linux脚本工具查找日志中的问题，为什么还需要日志易?**

一些公司的运维工程师在运维故障发生后，登陆各台服务器，使用grep/sed/awk等Linux脚本工具去日志里查找故障原因，排障时间长，未必能及时找到故障根源。日志易快速接收原始日志，统一管理并建立索引，能在几秒钟内返回搜索分析结果，帮助及时定位故障原因。

- **我们已经使用Hadoop处理日志，还需要日志易吗?**

Hadoop是个开发框架，用户需要开发Hadoop程序处理日志，使用门槛较高，优秀的Hadoop开发工程师不容易招到；

Hadoop是批处理，实时性较差：不少使用Hadoop处理日志的公司通常是每天晚上处理当天的日志，第二天出统计报表。有些公司做得好些但也只能看到几小时前的日志分析。

一些用户使用Hadoop框架下的Hive或Pig查询日志，延时可能达到几十分钟。

日志易采用流式实时处理框架，用户可看到十几秒钟之前的日志；而且日志易是拿来即用的Turn Key Solution，不需要用户做任何开发。使用日志易查询、分析日志，延时只有几秒。

- **我已经使用了开源日志分析系统，为什么还需要日志易?**

如果您的日志来源单一、日志量不大、对延时要求不高、有较强的开发能力，使用ELK等一些开源日志分析系统做一些定制化开发可能是可行的方案；但在日志来源多样、日志量大、对延时要求高、希望拿来就用的情况下，日志易是您的最好选择。

2. 产品概述

日志易是一款文本日志搜索引擎，通过功能强大日志分析功能，帮助用户及时发现问题。

日志易旨在为企业提供一个日志集中管理平台，提供一个配置简单、功能强大、容易使用的日志管理工具，通过对日志进行集中采集和准实时索引，提供搜索、分析、可视化和监控告警等功能，帮助企业进行线上业务实时监控、业务异常原因定位、业务日志数据统计分析、及安全与合规审计。

2.1. 服务模式

日志易提供在大中型企业内的本地部署版，也提供SaaS服务。

2.2. 产品特色

- 统一日志管理

日志实时采集，日志统一处理、统一保存、集中查看、集中搜索，无需登录每台服务器人工查看日志，提高运维效率，保障运维安全。

- 功能强大
    - 自动识别各种日志类型，自动抽取关键字段，将非结构化日志转化为结构化数据；
    - 对日志做全文索引，方便搜索；
    - 对不同来源的日志做关联分析，方便及时定位问题；
    - 丰富的统计及可视化功能，日志情况一目了然；
    - 对日志进行监控告警，可电邮通知用户；
    - 对日志、解析规则、分析模型等进行分组管理，赋予使用者不同的权限；
    - 提供开放API接口，灵活对接第三方系统或二次开发日志应用。
- 高性能、可扩展

采用高性能分布式架构，可支持每日TB级别的日志量，每秒钟可处理10万条日志，搜索分析结果秒级延时，让用户查询到十几秒钟之前产生的日志。日志易系统在普通PC服务器上运行，可根据日志量的增长逐步扩容。

- 操作简明易用

不需要用户掌握复杂的搜索语法，使用简单的搜索框查询，通过用户图形界面和友好的用户交互来完成强大的搜索分析功能。

3. 应用场景

1. 运维和应用性能监控（Application Performance Monitoring）
    - 通过日志对网络设备、服务器及应用程序状态实时监控，迅速定位问题根源
    - 通过日志对应用程序性能实时监控，及时发现性能瓶颈
    - 关联不同系统或模块的日志，进行端到端的服务监控和故障排查
2. 安全信息与事件管理（Security Information and Event Management）
    - 通过服务器日志发现端口扫描和非法入侵
    - 防火墙、网络设备、服务器日志安全跟踪分析
    - 信息安全合规审计
3. 业务统计分析
    - 网站用户及手机用户访问统计
    - 社交、视频、电商、游戏网站用户行为及交易行为分析
    - 客户端设备、操作系统、浏览器统计
4. 程序研发Bug分析
    - 快速关联分析大规模分布式系统各个模块产生的大量Debug日志

4. 产品简介

日志易构建的高性能、分布式日志处理架构可以每秒钟分析10万条日志，每天可以处理TB级的日志量，秒级处理延时，用户可搜索、分析几秒钟之前产生的日志。

4.1. 日志处理流程

- 日志采集：日志易使用Linux系统自带的rsyslog，或日志易agent，可采集服务器、网络设备、操作系统、应用系统的文本日志及二进制日志数据。支持标准的syslog协议（RFC5424）、HTTP和HTTPS。
- 字段提取：日志易系统通过系统自带或用户配置的规则引擎解析日志，抽取关键字段，将非结构化的日志转化为结构化数据。
- 路由索引：按路由规则将数据发送到指定索引中，并对关键字段及日志全文做分布式索引，方便用户对关键字段或全文进行检索。
- 备份恢复：可以同期发送数据到指定的备份路径中，并在对应的索引数据过期后手动恢复，继续查询使用。
- 搜索分析：像使用搜索引擎一样使用日志易进行日志搜索，查找满足特定条件的日志。
- 告警监控：根据用户预设的搜索告警条件自动检索日志，查看搜索结果是否满足预设告警条件，如果触发告警，则通过邮件或短信通知用户。
- 统计分析、数据可视化：对搜索结果进行统计分析和可视化，展示时间折线图、条形图、饼状图等，让数据分析更直观。

4.2. 功能介绍

4.2.1. 日志采集

- 日志易使用Linux系统自带的rsyslog或syslog-ng agent，或日志易agent，可采集服务器、网络设备、操作

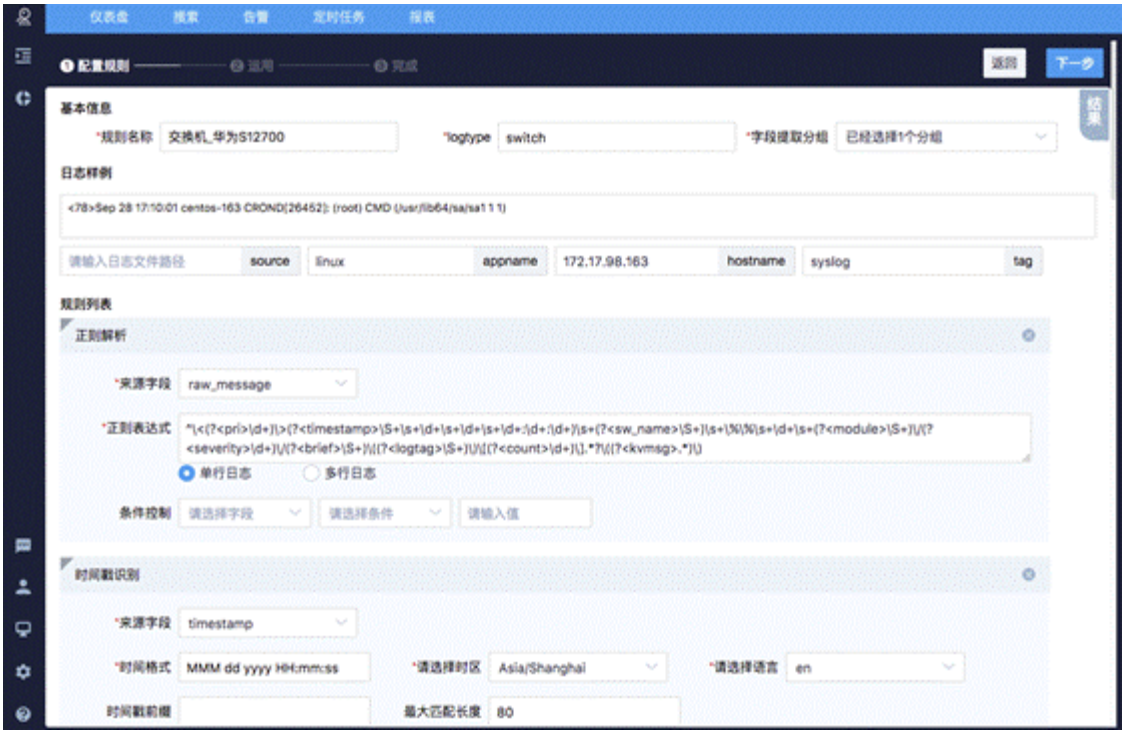
系统、应用系统的文本日志及二进制日志数据。我们支持标准的syslog协议（RFC5424）、HTTP和HTTPS。

- 向导式的数据采集流程

- 日志集中管理：无需登录单台服务器或授权开发人员访问生产环境，所有日志都可通过日志易Web界面授权访问。
- 支持各种类型的日志：任何基于文本类型的日志，无论来自服务器或是客户端，例如Apache、Java、PHP、Tomcat、MySQL、syslog-ng、rsyslog、nxlog、路由器等网络设备的日志，都可以上传到日志易。

4.2.2. 字段提取

- 自动解析日志：自动提取日志的关键字段，将非结构化日志转化为结构化数据，标准的日志格式支持有Apache、Nginx、Syslog、Java、JSON等。
- 日志格式自定义：为特殊日志格式量身定做，提供向导式解析规则配置，实现精准解析，提供正则匹配、KeyValue分解、url解码、时间戳识别、字典翻译、IP 地址库等多种提取方式简单操作，轻松上手。



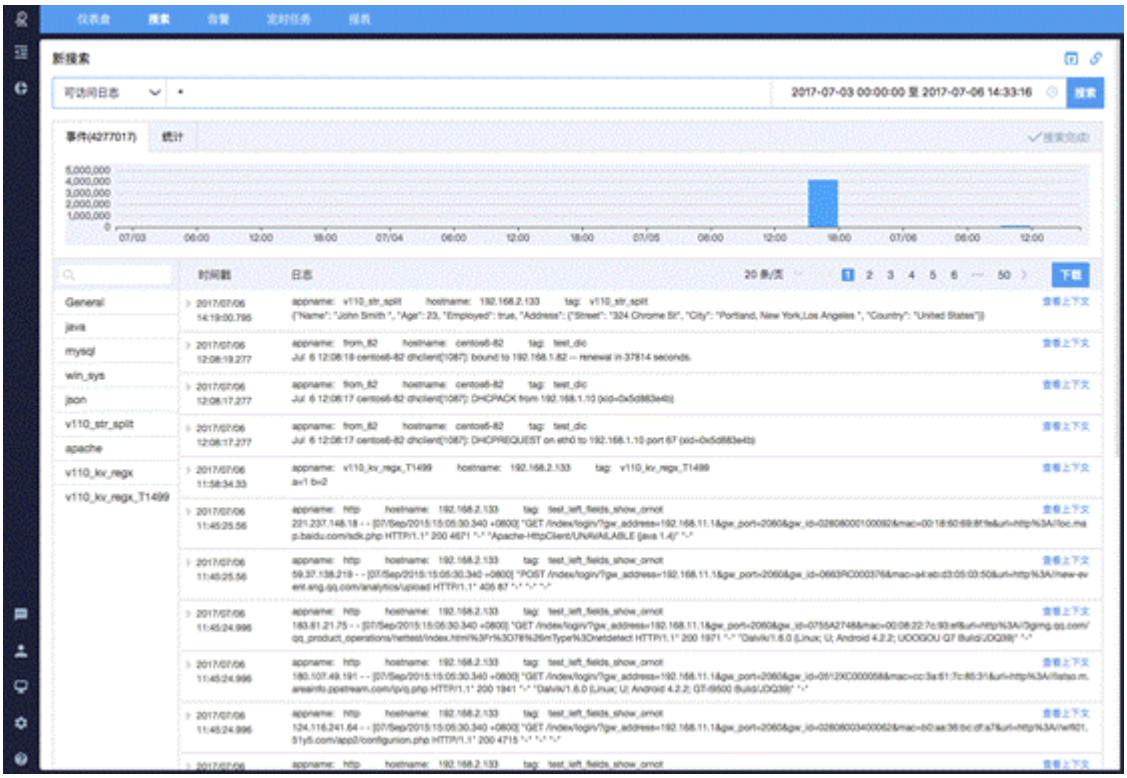
4.2.3. 权限管理

· 用户角色：建立完善的基于角色的用户权限体系。不同角色的用户使用日志易的功能及搜索、统计、告警等资源时，都可以限定到不同级别的权限；

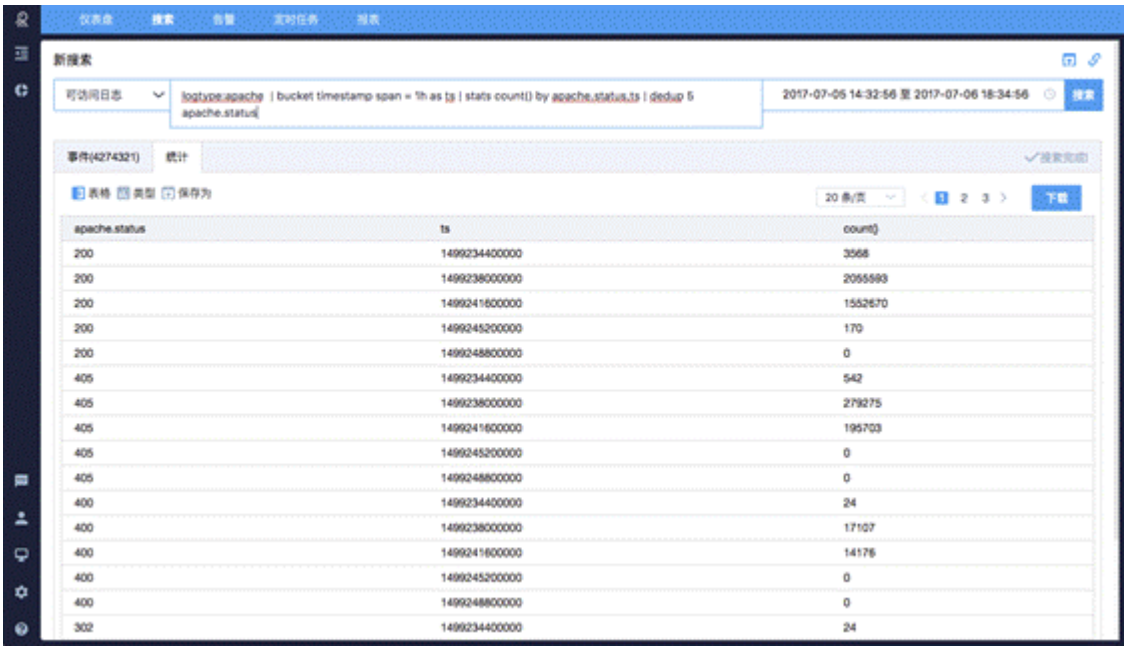
· 日志来源：用户可以动态定义感兴趣的日志来源，可根据主机、应用程序名、标签或自定义过滤条件创建日志来源，然后使用日志来源来限制搜索或分析特定范围。

4.2.4. 搜索

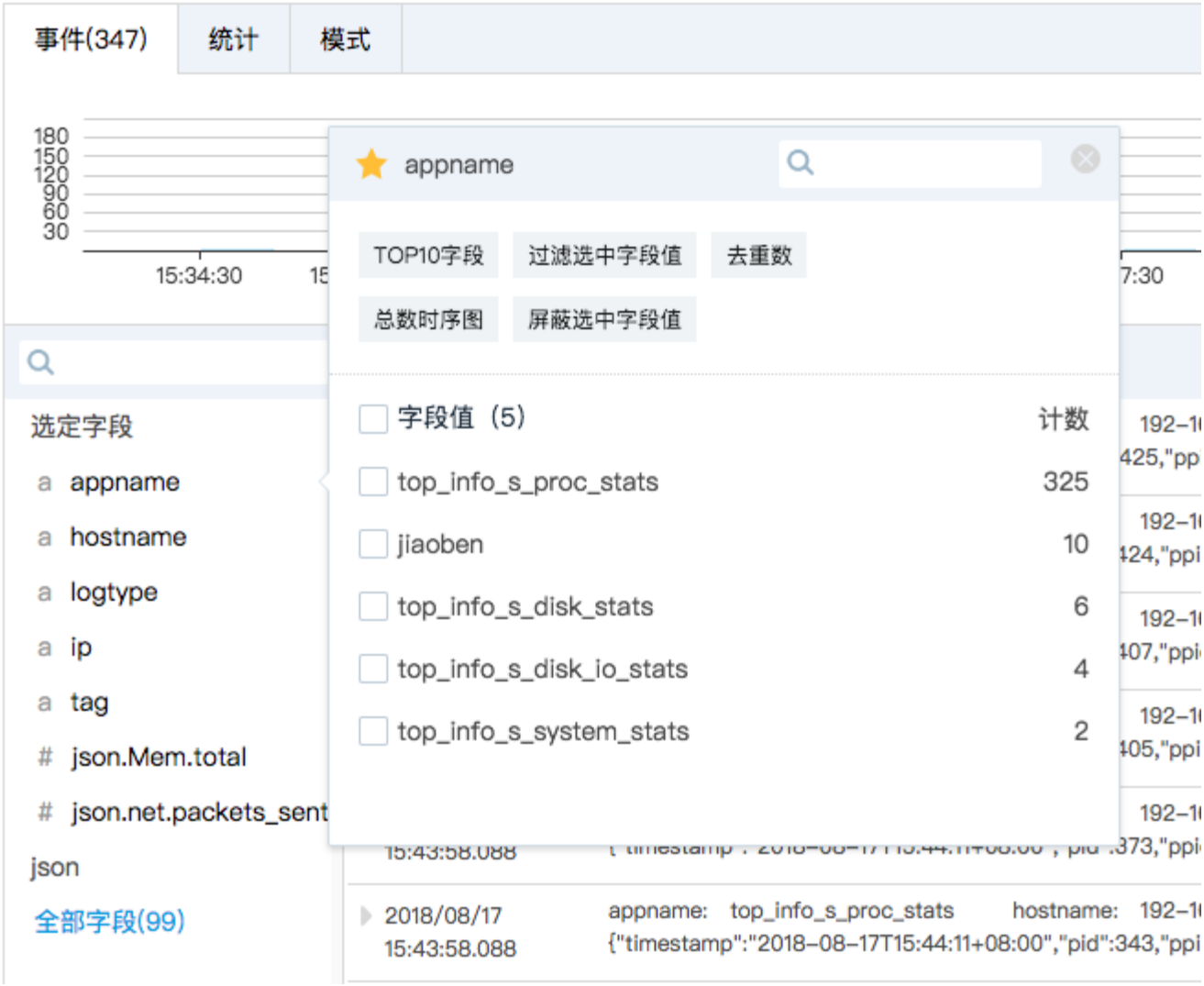
· 强大的搜索功能：支持全文检索，还可以使用字段、数值范围和布尔检索，查询指定时间范围内的日志。



· 完善SPL功能，支持新建字段、复杂的统计语句以及强大的关联搜索功能



· 字段值过滤：抽取日志关键字段后，用户可以点击展现的字段值实现日志过滤。



· 保存搜索结果，共享使用：所有搜索语句均可以保存，保存后可直接调用或共享给其他用户。

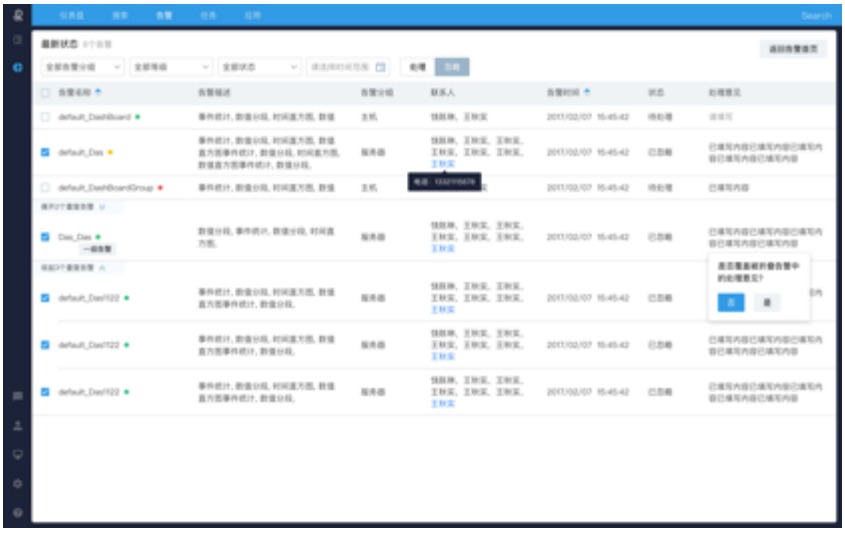
· 定时任务：通过设置搜索命令及任务执行时间实现日志检索自动化，任务数据长期留存，可供后续更高效的查询统计。也可以推送任务数据到第三方数据库中。

4.2.5. 监控

· 监控告警：基于已存搜索配置灵活的监控告警，支持秒级告警，当触发配置的告警条件时通过邮件、短信

等接口通知。

· 事件处理：对告警事件提供完善的归并、抑制、处理标记、事中操作功能，并提供处理状态统计报表。

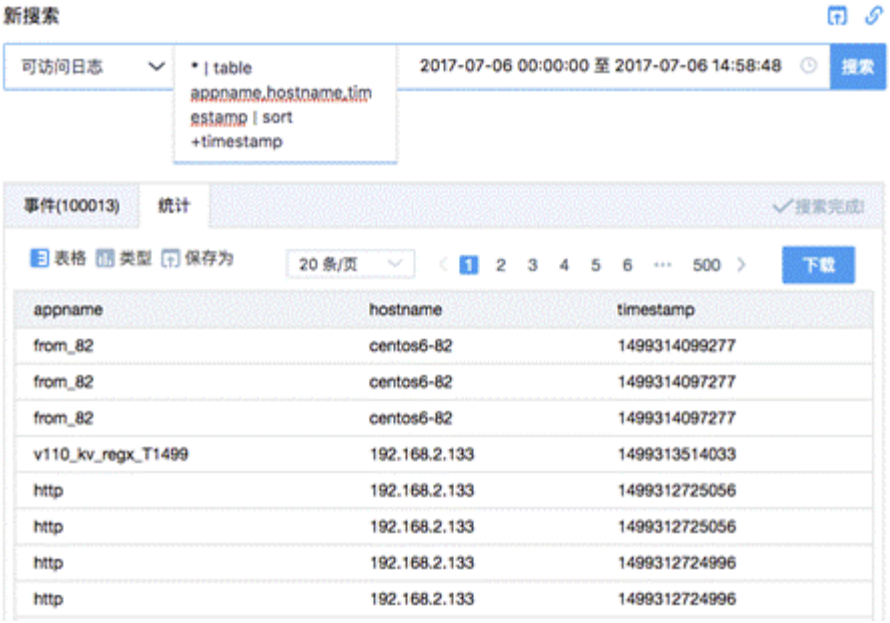


- 消息提醒：提醒内容包括告警消息、超量超限消息、未识别日志类型消息等，帮助用户第一时间掌握日志动态。

4.2.6. 可视化与统计

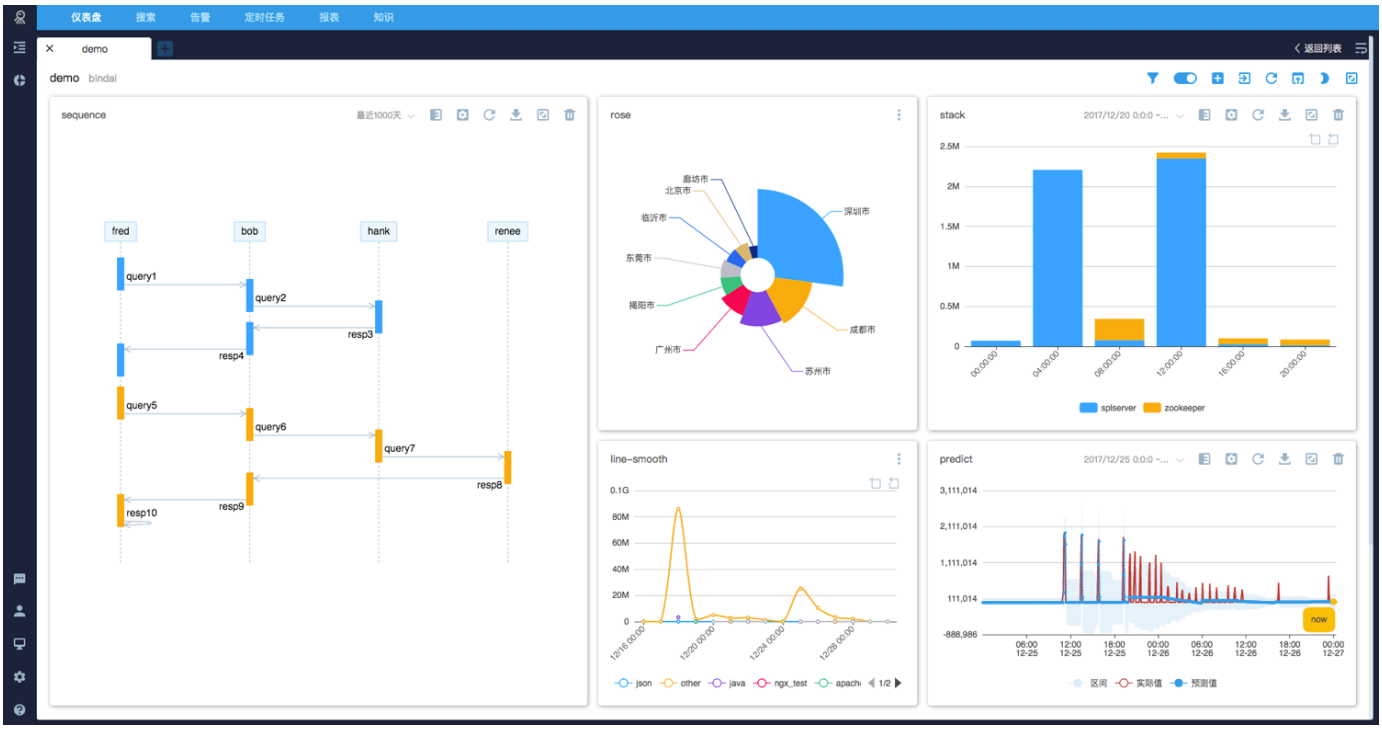
- 快捷统计：统计类型可支持计数统计、时间分段统计、数值分段统计、时间直方图、数值直方图、字段值分类、字段数值统计、累计百分比、多级统计、地图展示等多维度统计方式，数值型字段可从下拉菜单选择计数、求和、最大值、最小值、平均值、标准偏差等创建折线图、条形图、饼状图等不同形状的趋势分析图。

- 表格展现：结构化之后的日志数据可以通过类似电子表格的形式展示关键字段值，用户可以选择日志中感兴趣的关键字段，可以按单个字段列排序。



· 高级可视化：在电子表格的基础上，用户可以选择指定字段设置，生成自己想要的复杂可视化效果。日志易提供折线图、区域图、堆叠区域图、散点图、柱状图、分组柱状图、堆叠柱状图、饼图、玫瑰图、条带图、和弦图、桑基图、力引导图、区间图、多Y轴、区划图、热力图、轨迹图、单值、水球图、词云图、循序图等可视化效果。

· 仪表盘：通过趋势图组建仪表盘，拖放选择合适的大小来组织界面布局。仪表盘支持分组分享，支持自定义钻取和过滤功能，用户可以快速查看符合条件的数据。

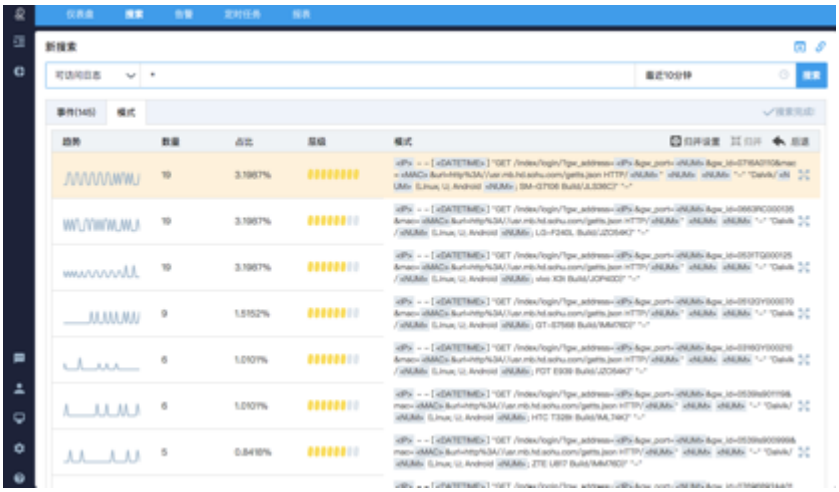


· 报表：拖拽方式自定义生成个性化日报、周报、月报，设置接收邮箱地址，方便查看图形化报表内容。

· 拓扑图：拖拽方式自定义业务拓扑和数据流转关系图，各拓扑节点支持按照业务关系分组，支持分区展示业务状态统计指标，支持点击钻取和定时刷新。

4.2.7. 人工智能

- 机器学习平台：日志易提供 20 多种标准机器学习算法，并支持在界面化交互平台上探索机器学习算法的数据运用。日志易按照数值预测、分类预测、离群检测、聚类分析、时序预测五个场景分类，提供不同的模型测试效果评价和可视化。
- 日志模式：日志易针对非结构化日志分析场景，提供专属的模式学习智能算法。方便用户快速了解百万行日志内含的行为模式和趋势，快速定位和发现未知故障。



4.2.8. 集成

- API：日志易提供对日志易系统的 RESTful 调用接口，当前仅作为日志易企业版的一部分，访问 RESTful API 需要进行签名验证，便于用户以灵活的方式集成日志易系统。
- SDK: 日志易提供 Java、Python、C# 版本的 SDK 库，便于用户程序集成日志易系统。

- Grafana: 日志易提供 Grafana 专属的日志易 Datasource 扩展插件。用户可以在 Grafana 大屏上无缝集成日志易查询和数据展示。

5. 系统环境要求

5.1. 服务器

- 硬件： - x86架构PC服务器（详见硬件推荐）
- 操作系统： - CentOS 6.5 X64服务器版本
- CPU： - 至少4核（根据日志量而定）
- 内存： - 至少8GB ,建议16G或以上（根据日志量而定）
- 硬盘： - 至少300GB（根据日志量而定），推荐RAID 0+1

5.2. 客户端

Web浏览器： - 所有常用浏览器，包括Chrome，Firefox，Safari，IE10以上版本，360浏览器，搜狗浏览器等

如果想获得比较好的体验，建议:

- Chrome45+
- Firefox40+
- IE11
- Edge浏览器