DBAplus

云与运维技术沙龙-广州站

企业运维大数据日志分析实践

刘岳东

目录

DBAplus

- 我来干什么
- 我们怎么干
- 如何证明

日志系统与传统网管的区别

DBAplus

传统网管告警

Gigabit[illegible]状态为Down

网管用来发现问题

现有网管告警，根据设定阈值告警，告警简单、信息单一

日志分析告警

Jun 11 23:59:59 188.9.11.124 VSYS=public; AttackType: Trace route attack[illegible]cpu: 3;
Receive Interface: GigabitEthernet3/0/0 GigabitEthernet1/0/0: f[illegible]97
221.177.225.167 117.142.2.71 223.103.1.39 117.142.7[illegible]177.142.201
221.177.100.214; begin time: 2016-06-11[illegible]

检索字段：

- [illegible]23:59:59
- d[illegible].9.11.124
- attacktype = Trace route attack
- slot: 5;
- cpu: 3;
- Receive Interface: Giga[illegible]
- fromip = [illegible].142.2.71 223.103.1.39
  1[illegible]
- to[illegible]221.177.100.214
- beg[illegible]= 2016-06-11T15:59:29.000Z
- endtime = 2016-06-11T15:59:57.000Z
- appname = /root/netlog/hw.log.bak1
- hostname = shsnc-04

日志用来定位问题

日志用来预测问题

日志可以提供非常丰富的上下文信息，帮助人员快速定位故障

并结合传统网管告警形成更加完整的告警检索体系

运维分析体系建设的四个阶段

DBAplus

PART ONE
0

1

基础管理

基础设备监控
智能告警
ITIL流程

PART TWO
0

2

解决方案与工具

企业资产配置与管理
性能管理
容量管理
故障管理
自动化运维
日志分析

PART THREE
0

3

平台整合

SAAS
PAAS
Cloud化
大数据分析
综合网管
管理制度化

PART FOUR
0

4

以业务维度为核心

业务感知
关联分析
闭环管理
持续改进与优化
人工智能

日志的需求

日志的分类

- 01 IT基础架构日志
- 02 平台/软件日志
- 03 业务分析日志

日志的特点

以前是怎么干的

DBAplus

日志来源	基础设施	软件平台	操作系统	数据库/中间件	业务系统
	• 机柜设备 • 服务器设备 • 网络设备 • 存储设备	• 工作流系统 • 电子表单系统 • 报表软件 • 自动运维系统	• UNIX系统 • Linux系统 • Windows系统	• 关系型数据库 • 消息中间件 • Hadoop • Kafka/JMS	• Log4j • 自定义日志 • 业务数据
	• 组件温度日志 • 风扇异常日志 • 电源模块日志	• 日常运行日志 • 异常告警日志 • 错误故障日志	• 资源使用日志 • 系统事件日志 • 操作运行日志	• 操作运行日志 • 访问监听日志 • 错误故障日志	• 异常诊断日志 • 业务时序日志 • 用户体验日志

日志数据膨胀速度快，最后简单丢弃
分散存放，碎片化难以管理
日志数据格式不一，难以分析
日志数据分析单一，没有关联
对日志数据的安全分析比较简单
缺乏直观的报表，无法挖掘有价值的信息
无法实时告警，无法进行全面的合规审计

现在有些企业是怎么干的

DBAplus

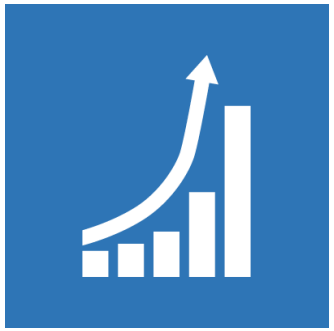
Elasticsearch

- ✓ 实时检索、亿级数据秒级返回结果
- ✓ 分布式系统，对外表现对等，自动均衡
- ✓ 输入/输出 JSON
- ✓ 多租户，不同的用途分索引
- ✓ 可以同时操作多个索引
- ✓ 基于Apache Lucene搜索引擎库

Logstash

- ✓ 分布式集群架构、模块化设计
- ✓ 安装部署方便，支持不同操作系统
- ✓ 扩展性强，支持自定义插件，内置的120多个正则格式
- ✓ output插件，自带90多种插件
- ✓ 性能配置性强，通过配置方式组合各类插件
- ✓ 支持不同的输入源及输出，自带的各类input、filter
- ✓ 数据处理灵活性强

Kibana



- ✓ 报表配置灵活，易于共享
- ✓ 集成丰富的图表库, 支持复杂的分析报表
- ✓ 自由展示报表大小及位置
- ✓ 数据的实时统计分析
- ✓ 与Elasticsearch的无缝集成

现在有些企业是怎么干的

DBAplus

与开源ELK对比

无资源控制

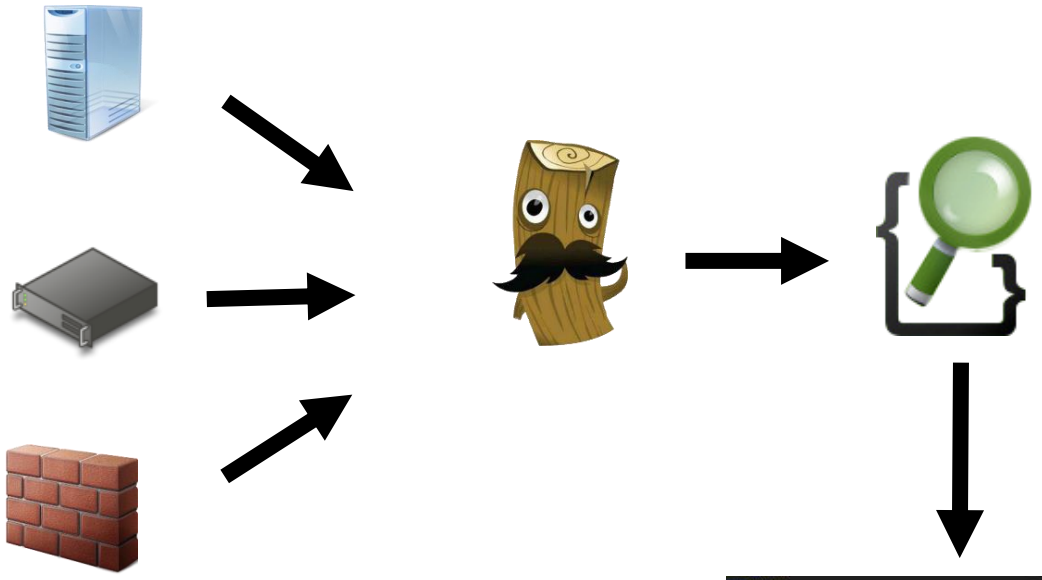
没有消息队列缓存

告警功能弱

用户认证及权限管理

不支持关联统计分析

无自身监控能力



无商业化支持

无机器学习

无运维场景

无行业经验积累

难以定制性开发

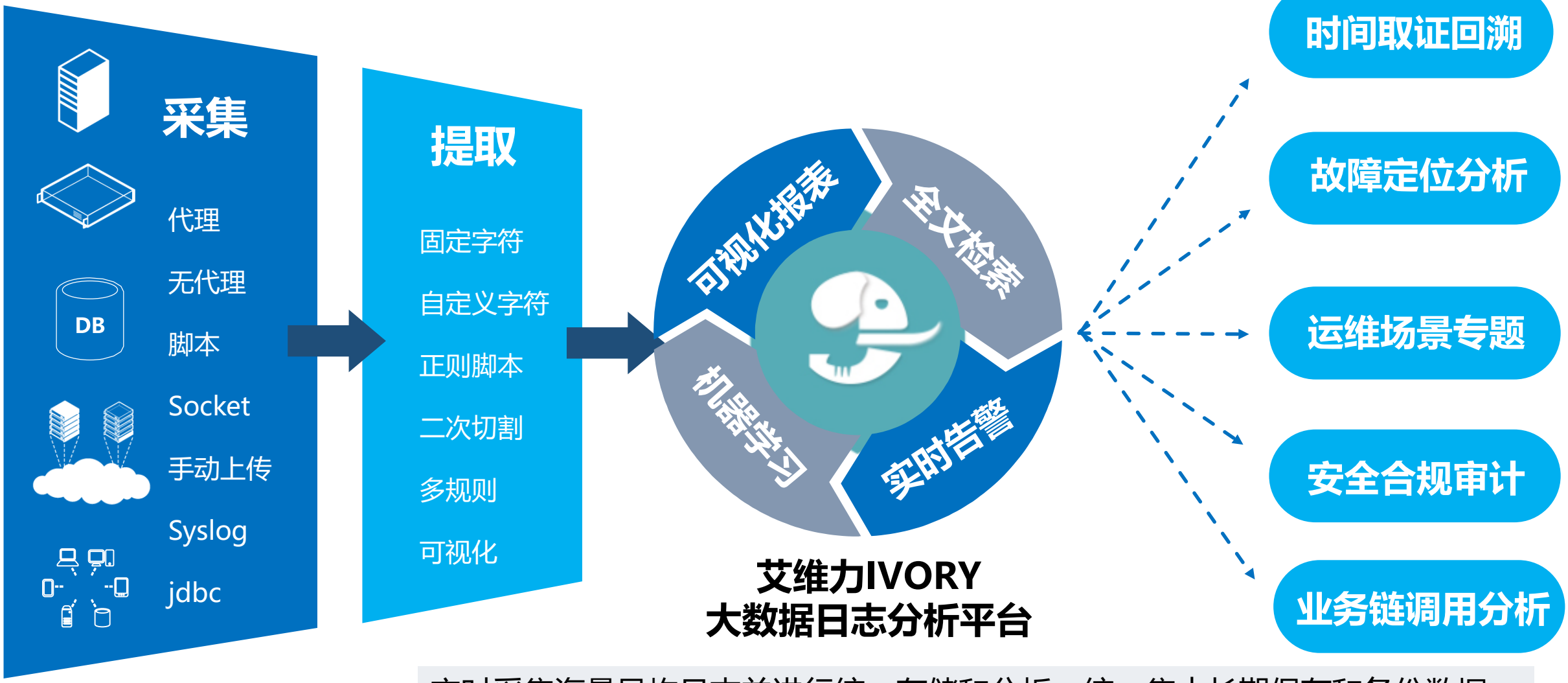
无数据脱敏

无法即开即用

类似开源软件hadoop，越往后维护的成本越高

我们是怎么干的

DBAplus



实时采集海量异构日志并进行统一存储和分析，统一集中长期保存和备份数据，发现系统和业务中潜在的问题隐患，提升运维分析和故障处理效率

内置识别算法，数据智能规范化

DBAplus

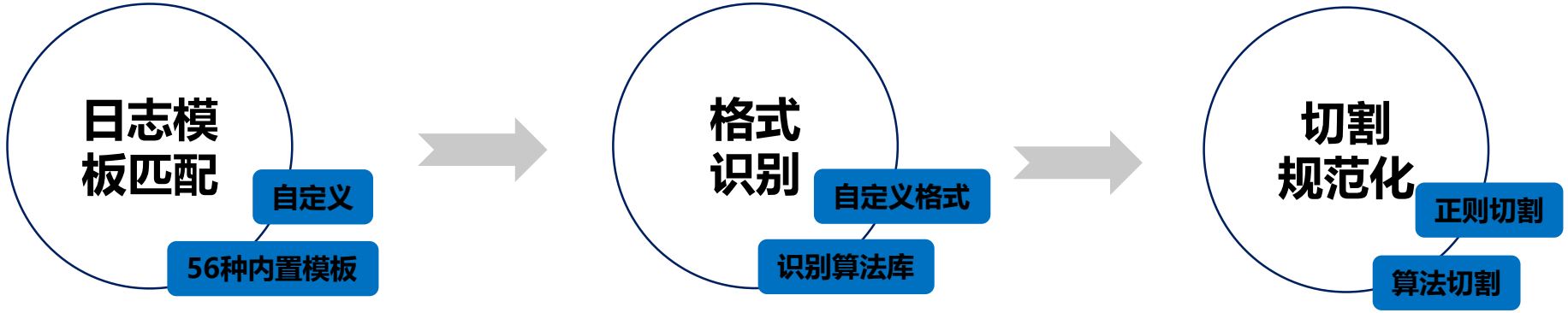
数据生产源

网络设施日志数据	虚拟化平台日志数据	操作系统日志数据	数据库/中间件日志数据	业务系统日志数据

原始IT日志

217.227.233.68 - - [13/Sep/2012:02:38:26 -0400] "GET /images/jordan-80.png HTTP/1.1" 200 6146 "/" "Mozilla/5.0 (X11; Linux x86_64; rv:15.0) Gecko/20100101 Firefox/15.0"http://www.semicomplete.com/articles/ssh-security

识别库



格式化

Clientip	Useragent	Status	Req_time	Version	Method	Referer	URI_Path	Bytes
217.227.233.68	"Mozilla/5.0 (X11; Linux x86_64; rv:15.0) Gecko/20100101 Firefox/15.0	200	13/Sep/2012:02:38:26 -0400	HTTP/1.1	Get	http://www.semicomplete.com/articles/ssh-security	/images/jordan-80.png	6146

识别度转换

时间格式：

- 020805 13:51:24
- [2013-07-09 15:49:44,385]
- 29/Apr/2011:07:05:26 +0000
- Fri, 21 Nov 1997 09:55:06 -0600

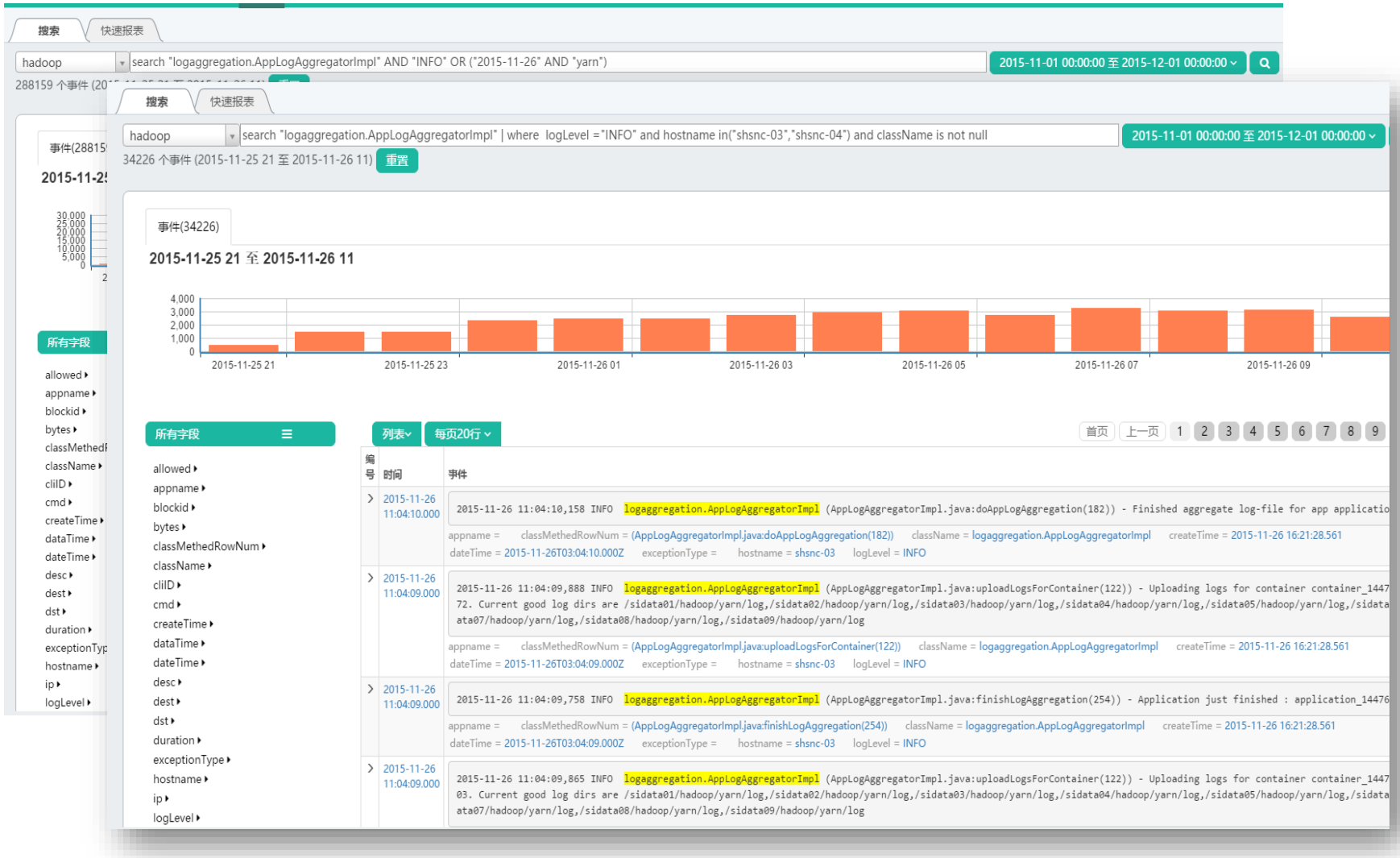
- 2012-09-13 02:38:26-0400

兼顾非结构化和数据库格式

机器对文本的读写性能远胜于传统数据库，但是数据库格式化的数据才能真正实现商务智能分析

搜索引擎上的函数语言

DBAplus



逻辑搜索检索

- ✓ 系统支持输入 “and “, “or “ 逻辑关联运算搜索
- ✓ 增强分析的逻辑性

逻辑范围运算检索

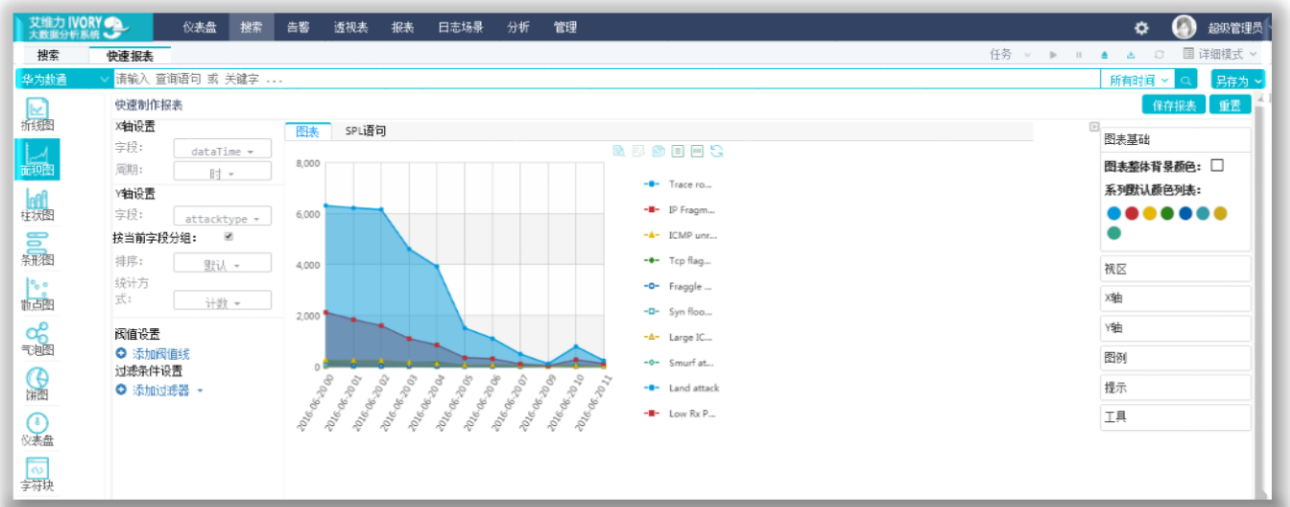
- ✓ 系统支持输入” <,>,<=,>=,+,-,in” 等等逻辑运算的

函数检索，算法支持

- ✓ 支持聚合函数(count,sum)，字符串/数字/时间操作函数；可支持强大的分析统计功能
- ✓ 在线支持算法如关联算法，关联分析等

类SQL搜索语言使一般用户也能快速熟悉分析语言，提供在线统计分析、算法、挖掘功能，进行各种问题回朔和取证

自助分析报告：强大的报表自定义能力，轻松实现简单数据分析



报表自定义、大数据在线分析

- ✓ 无需透过其他工具直接产出报表，支持双维和多维分析
- ✓ 支持多种运算函数：计数，去重，统和等等，强化报表分析报告准确性

丰富的图表库支持

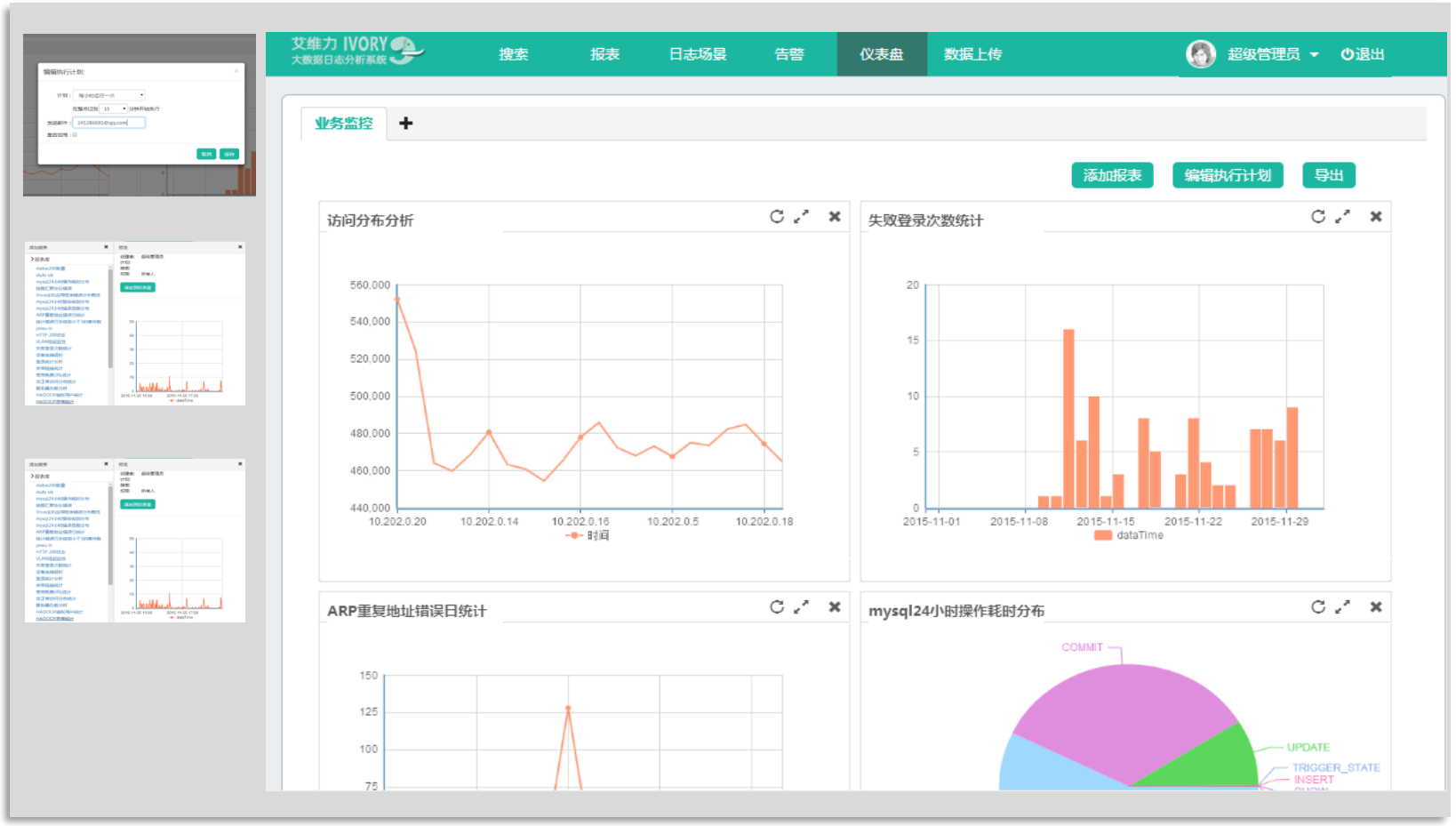
- ✓ 支持12种常规报表模式，如：二维报表，多维报表，饼状图，趋势图，直方图，分区图，散点图等等

场景预演和沉淀

- ✓ 生成的报表可以直接转换为存储到仪表盘，沉淀成场景
- ✓ 内置定时器和调度机制，数据自动刷新

无需任何技术背景，更可弹性化地产生企业管理阶层所想要的报告内容。

仪表盘功能：运维场景自定义，提升订阅式自助运维能力



01 灵活、任意装卸

- ✓ 根据分析思路沉淀场景，在仪表盘可以任意灵活插拔分析报告

02 定时调度

- ✓ 每个分析报告都有独立的执行计划运行规律和周期型获取数据变化
- ✓ 平台支撑超过百个任务同时调度

03 自动交互

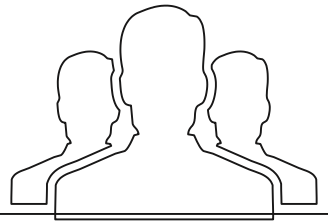
- ✓ 每个分析报告都会根据定时调度任务进行异步交互，自动刷新数据更新

分析思路沉淀和固话成为场景，通过图表趋势/分布一目了然的呈现在仪表盘

运维视角时间轴分析

DBAplus

场景分类



运维专家总结

IVORY场景分类

故障类

- 异常URL统计分析
- 网络异常根源定位
- 系统异常关键链
- 错误页面异常定位
- 磁盘坏道根源定位
- 硬件问题诊断定位
- ORACLE实例挂起定位
- ……

安全类

- 账户暴力破解根因
- 访问地址来源统计分析
- DDoS攻击分类分析
- 服务验证失败类别统计
- 系统su状态会话统计
- 系统未知限制的类别分布
- ……

隐患检测

- 主备切换异常隐患发现
- 数据备份潜在隐患发现
- 系统后门漏洞发现
- 网络使用率
- Timeout 使用率
- 日志量异常
- ……

性能类

- 页面访问缓慢分析
- 系统资源紧张分析
- 文件读写缓慢分析
- JAVA内存溢出分析
- 死锁和线程问题分析
- ……

审计类

- WINDOWS系统事件
- Linux日志审计
- 中间件访问审计
- 移动设备操作审计
- 应用崩溃审计
- ……

业务分析

- 网络访问兴趣分析
- 金融风控
- 电商充值
- 交易量、失败率分析
- 用户行为分析
- ……

部分场景展示

DBAplus

路由配置报表	路由器或交换机系统事件	路由器登录报表
链路连接	命令执行	登录
链路断开	接口连接	失败的登录
链路连接和断开	链路失败导致接口关闭	错误的身份验证
链路状态更改	单点端口停止	SSH登录
配置更改	风扇失败	SSH失败登录

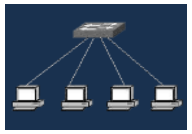
多层业务关联分析	监听分析	登入登出	Client连接情况
操作发生的URL	泄密通路回溯	数据库错误	Hight TRC分析
客户端的IP、请求报文	数据库风	SQL语义分析	Alter System分析
SQL命令排行		操作时间分析	DB对象分析
管理风险展示	操作	越权操作	共用账号展示
弱鉴权机制	网络层攻击统计	漏洞统计	SQL注入查询
拒绝服务攻击	应用程序攻击	可以端口分析	可以进程分析
日志完整性	性能查询		

数据库场景

路由器流量错误	基于ICMP流量审计排名	基于用户的SSH失败登录排名
过多分片	基于流量审计排名	登录尝试
无效分片长度		失败的登录尝试
重复分片		

大数据日志分析体系建设步骤

DBAplus



01 日志分析平台

集中化日志分析平台

- 分布式采集
- 统一存储
- 实时检索
- 报表分析
- 告警推送
- 大屏展示

02 大数据运维分析平台

提供其他平台数据接入，对系统和业务异构数据进行整合和分析，提供更多的场景：

- 调度管理系统
- 关联分析
- 元数据管理
- 数据安全系统
- 可视化分析系统

03 智能运营平台

基于机器学习算法构建智能分析系统，基于业务为企业提供预测和决策支持

- 个性化推荐系统
- 舆情洞察系统
- 用户标签管理系统
- 问题预测

怎么证明-案例分享

DBAplus

中国移动 China Mobile

广东　浙江

山东　四川

中国电信 CHINA TELECOM

湖北　湖南

博时基金

景顺长城 Invesco Great Wall

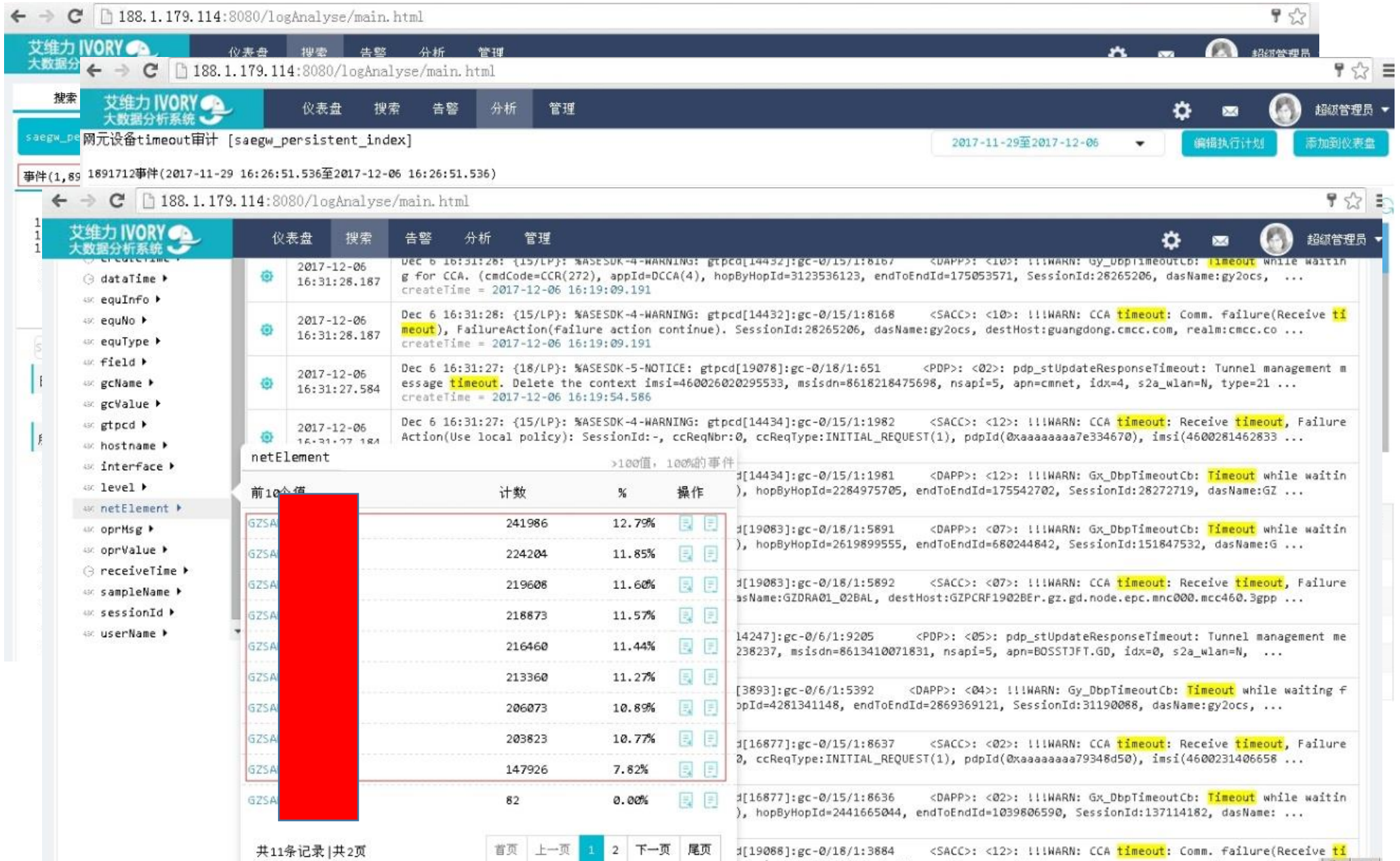
广东华兴银行 GUANGDONG HUAXING BANK

安吉汽车物流有限公司 ANJI AUTOMOTIVE LOGISTICS CO.,LTD

服务平台的广泛应用

IVORY日志大数据分析平台，已在多家大型企业使用，主要涉及电信、金融、交通、政府等领域，典型客户包括博时基金、景顺长城基金管理有限公司、湖南电信、山东移动、广东移动等。客户每天采集处理入库的日志量最高达3.8T，实现T级别日志数据查找秒级响应的速度，并为客户提供第三方日志分析专家顾问服务

案例分享-XX移动



1、监控最近7天数据有180万件timeout事件

2、在网元维度对timeout事件进行分布分析：timeout事件主要集中在9台网元设备上。

3、检索timeout事件网元占比

案例分享-XX电信

DBAplus

进一步分析表明，99.99%的攻击来自一个土耳其的IP 188.x.xx.xxx，建议对该IP进行处理

通过分析日志，发下如下两个源地址特别频繁，建议通过抓包确认下：

117.xx.xxx.xx:80

117.xx.xxx. xx:80

华为防火墙一共受到接近22万次

约18万条为
rst-flood攻击，
还有Smurf、
ragment、
ggle 、 Ip
ion source
te attack和
ge ICMP
ack

案例分享-XX电信

1、华三交换机日志在12月6日13点出现日志暴长。日志量达1800，为平时正常情况下的3倍

2、排查该时间段内的日志记录，发现在13点这一个小时内从同一个IP地址发起登入、登出操作达900次。

3、查看设备IP分布，在**10.201.xx.xxx**及xxx两台机子上都有出现

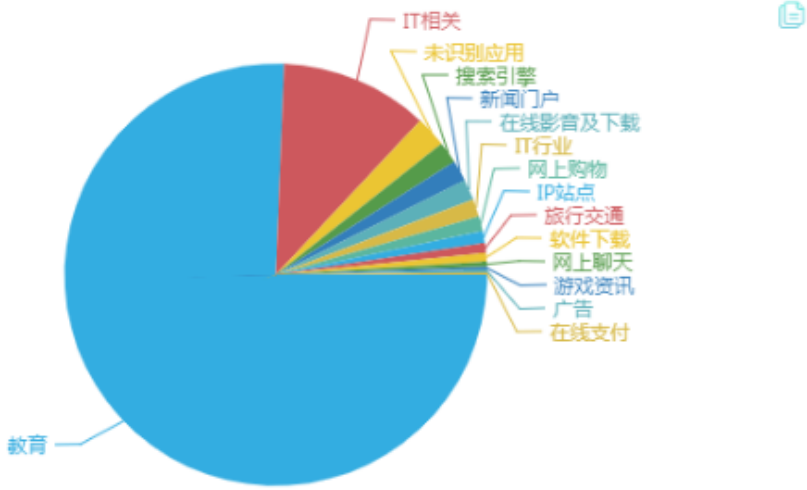
案例分享-XX高校

DBAplus

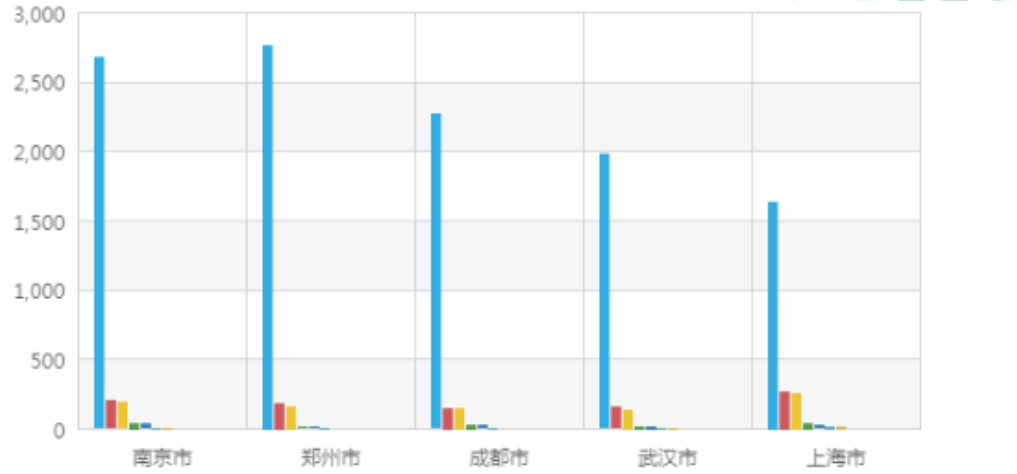
XX学院接入学生上网行为数据，利用Ivory分析系统建立上网行为专题，实时掌握学生上网行为，对异常行为内容进行重点关注。

开启更新 0分钟 添加报表 编辑执行计划 共享

网站分类访问最多的top15排序



链点请求不存在(状态码404)



最常访问网站TOP 15

app	count的计数
教育	207634
IT相关	31181
未识别应用	6483
搜索引擎	4554
新闻门户	4441

访问教育分类网站最多的TOP10学生

user	count的计数
201641101110	14366
201441404204	12590
201442302131	10178
201641302211	7891
201641101516	7323

我是谁

DBAplus

新炬网络
SNC Net

能力

资质

本地化

艾维力 IVORY
大数据分析系统

DBAplus

www.dbaplus.cn

THANK YOU !