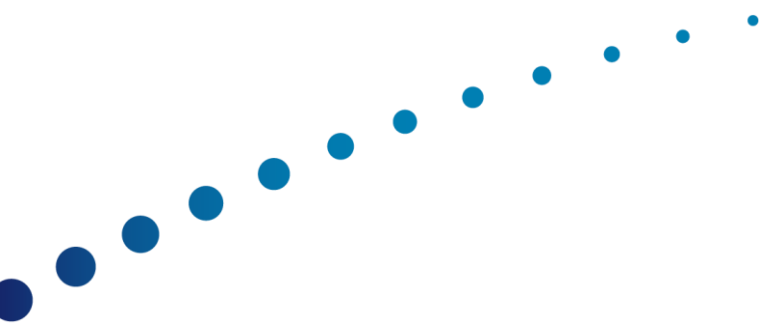
Auditsys4.3 产品改进说明

北京华夏博格数据技术有限公司
2019 年 6 月 27 日



修订记录

版本	变更描述	变更人	日期
1.0	创建		2019.6.27

目　　录

1. 4.3 版本规划与定位

4.2 版本主要完成了 Auditsys 从 windows 架构到 linux 架构的转换，数据库引擎从 Sqlserver 关系数据库到 Elasticsearch 数据库的转换。但是在 Agent 的探针并不多，Center 的数据展示分析能力较弱。

4.3 版本主要规划目标

1. 在业务层面

agent 端通过丰富的各类探针(数据库、邮件、qq、工作时长等)来强化 Auditsys 审计和分析功能，

center 端提供更多的图形化性图表和功能来优化管理员获取风险、数据、分析结果的能力，提高用户的使用体验。

2. 技术层面。

增加 linux 的客户端，并且在 server 中增加 linux 通道，实现对 linux 的监控，并为未来的 linux 桌面监控打下基础。

在 server 端增加敏感词分析模块，可以对用户的剪贴板、邮件等进行更细粒度的监控。

在 agent 与 server 端增加文件通道，用于文件拷贝、U 盘拷贝等文件内容，以及后续规划的邮件附件内容等文件内容识别及敏感词分析。

在 center 端增加 linux 与 windows 播放器融合，可自定义指标、图表、报表等。

在 ES 端把风险数据、敏感数据、工作效率数据等独立出来，方便多风险和记录的汇总聚合计算。

2. 优化改进

2.1 Agent 优化

2.1.1 *数据库探针*

监控客户端操作数据库行为,并记录数据库操作指令。目前能支持 navicate、plsql、sqlserver 数据库工具，能支持的记录指令的数据库有 sqlserver、oracle。

技术原理：通过技术手段监视数据库工具，监测用户的操作行为并截取的数据流报文分析并提取。缺陷：因不是直接读取的方式，需要针对每种数据库工具、每个版本单独适配抓取数据。

2.1.2 *邮件探针*

监控客户端操作邮件工具发送邮件的内容，并记录邮件的发件人、收件人、抄送人、密送人、标题、内容、附件名称。目前支持 OUTLOOK、foxmail 邮件工具监控。

技术原理：通过技术手段监视截取邮件工具，在点击发送时获取邮件的数据报文分析并提取。

2.1.3 *QQ 探针*

监控 QQ 客户端操作监控发送聊天内容，并记录聊天对象，发送的聊天内容记录。目前支持最新版的腾讯 QQ。

技术原理：通过技术手段监视截取腾讯 QQ 工具的窗口，在发送时获取发送的数据报文分析并提取。

2.1.4 *文件操作探针*

监控系统文件操作行为，记录用户对文件或文件夹操作的操作类型，操作文件夹名称，目的位置名称，文件数，文件大小。

技术原理：通过技术手段监视系统文件操作，在操作完成后记录相关信息。

2.1.5 *键盘鼠标事件探针*

监控用户的键盘鼠标行为，记录用户的键盘按键行为和鼠标按键行为。

技术原理：通过技术手段监视系统的键盘、鼠标操作行为，在操作完成后记录相关信息。

2.1.6 *USB 事件探针*

监控设备的 USB 事件（主要是 U 盘）行为，记录用户在设备上的 USB（U 盘）插入、弹出事件。

技术原理：通过技术手段监视系统的 USB 操作行为，在操作完成后记录相关信息。

2.1.7 *操作标签探针*

操作标签探针主要针对于打开应用或打开网页上，点击了某个按钮的行为识别。因某些操作既不符合应用记录（记录窗口切换），又不符合 web 记录的规则（记录网页地址切换），但我们需要监控的敏感行为。其中比较典型的两种：1）在具体的业务系统上点击了“上传”按钮　2）在某个应用窗口内，点击查看安装过程，但是没有点击“确认”修改。

探针会记录任意窗口和网页内的左键点击行为，并记录鼠标点击位置的内容，以及窗口标题、exe 名称、

技术原理： 通过技术手段监控鼠标左键，并且识别左键点击的位置，判断点击的位置是否是按钮、链接、文本等。如果是应用窗口，则记录应用信息+左键点击位置的信息；如果是浏览器，则记录浏览器应用信息+左键点击的位置信息+当前 URL 地址。

2.1.8 *工作效率探针*

工作效率探针主要针对于统计使用应用或打开的网址的时长。用于分析用户的使用工时，并根据各种策略计算用户的工作效率。

技术原理：用户在打开一个应用窗口后开始记录开始时间，用户点击切换会判断是否还在该应用，若在该应用则继续保持，若更换应用则结束上次的应用时长记录。或者停留 60s 不动后会将将该应用创建结束（该 60s 时长可以通过 center 界面中记录策略内配置）。若是浏览器应用，系统不会记录浏览器应用时长，而是单独记录打开的网址的时长，规则同应用。

2.1.9 *文件传输通道*

针对 U 盘的文件外拷操作，剪贴板操作，邮件外发操作将数据以附件形式外发的行为，截获附件并将附件上传到 server 进行敏感词分析的文件传输通道。（目前开放支持文件拷贝到 U 盘文件上传）

技术原理：监控用户拷贝或剪切文件到 U 盘，触发文件上传操作。如果是拷贝，则从源目录上传数据到 server，如果是剪切，则从 U 盘的目的目录上传到 server。备注：若剪切数据到 U 盘，若 agent 没有来的及上传到 server 就拔出了 U 盘，会存在部分数据无法分析到的情况。离线剪切不会记录上传，离线拷贝会从源目录上传。

2.1.10 *访问控制*

后台可以配置，阻断用户打开某些风险应用或风险网址。

技术原理：其中阻断应用的方式为杀死风险应用进程，阻断风险网址为强制跳转到指定页面。

2.1.11 *水印功能*

水印功能主要用于为防止外部人员通过截屏手段泄露数据。目前支持桌面水印和应用水印。

技术原理：通过在桌面或应用上覆盖图层实现。

2.1.12 *linux server 探针*

监控 linux server 的用户操作行为。目前支持 centos、redhat 6，7

技术原理：记录用户登录到 linux 的操作及回显（包括用远程工具登录的操作）。

2.1.13 *网页内容探针*

网页内容探针主要针对于保险行业中可能在 b/s 业务系统中可能有隐藏的敏感内容的挖掘需求。（该功能还在技术预研中）

技术原理：用户每打个一个网页，通过技术手段爬取网页报文内容，并识别敏感词数据。

2.2 Server 优化

2.2.1 *linux 数据通道*

增加了 linux 的 agent 数据接收，linux 的操作数据并没有录屏数据。linux 的操作数据是以元数据的形式存储到 ES 中。

2.2.2 *文件通道*

增加了与 agent 的文件通道对接，并增加了文件（word、txt、ppt、excel、pdf）内容的识别、转换。并将命中的敏感的片段存入 ES 敏感词数据内。

2.2.3 *风险规则优化*

优化风险规则，支持一条记录匹配多个风险规则，风险规则增加数据集匹配。

2.2.4 *敏感词功能*

增加独立分析模块敏感词分析。因触发敏感词的内容都是大文件或长内容，正则匹配效率较低，目前敏感词分析支持正则匹配手机号、身份证，以及自定义数据集合。目前已接入敏感词分析的数据源有几种：邮件、QQ、文件操作、剪贴板、网页敏感内容。

2.2.5 *工作效率分析计算和匹配*

agent 记录应用工时和 web 工时后，在 server 中计算每条记录的上班时长和加班时长（根据 center 端配置）。并且会根据在 center 中配置的工作效率规则，对记录进行标记分类。

2.2.6 *服务器归类*

服务器可以根据 IP 段或组织进行分配。

2.2.7 *自定义终端数据对应融合*

存在部分客户需要直接对终端进行部门和用户姓名的标记（招商证券、广州公积金）。在终端设置好后，将数据融合到用户产生的数据中，并可以聚合分析。

2.2.8 *替换 Kafka*

从 4.2 版本以及使用的效果来看，使用开源的 kafka 存在重启，消费数据丢失等问题。再 4.3 中已逐步移除 kafka，而使用自己开发的数据缓存及读取标记功能替代，类似于简化版的 kafka 的功能设计，在数据压测和稳定性方面更好。数据缓存默认在/home 目录下 sque , squepos 文件夹。

2.3 控制台优化

2.3.1 *播放器改造*

增加了 linux 客户端内容的数据播放，采用 termail 方式展示。和原有的 windows 操作集成播放，并且增加在每帧内可以点击参考帧详情。

2.3.2 *视频下载*

目前支持 html 格式的视频文件包下载并播放。

2.3.3 *行为数据接入*

增加了数据库、USB、邮件、QQ、文件操作等数据的检索功能。

增加了模糊搜索功能。

2.3.4 *更灵活的管理员授权*

针对不同的客户可能对审计管理员的数据查看权限有不同的要求。之前统一按照组织进行授权，现已优化为可按组织、部门（通过终端栏目自定义配置）、用户组三个维度组合配置。

2.3.5 *终端关系功能*

部分客户需要在终端中配置对应部门、用户的姓名、扩展字段。并且通过配置的字段进行数据过滤，管理员权限的过滤等。

提供手工编辑、excel 导入、联软导入三种途径配置对应的数据。

2.3.6 *终端无会话和服务器容量告警*

针对终端长时间没有会话和服务器视频存储目录空间不足的问题，提供单独的告警功能。

2.3.7 *数据集功能*

在风险规则以及敏感词规则的条件匹配中，存在某个条件需要匹配很多的选项，直接写在风险规则中修改和查看都不太友好，所以提供统一的数据集合功能，把需要匹配的项归集到一个数据集中，在规则配置中，只需要配置在...中即可。

2.3.8 *工作效率功能*

工作效率配置：提供工作效率上班和加班配置，默认周一到周五 9 点到 17 点为工作日，其他时长的数据为加班。提供例外工作日和例外加班日的自定义。

工作效率分类：可以通过自定义配置方式标记用户行为的工时数据分类标签。

工作效率明细：提供查询工作效率明细查询。

工作效率分析：提供自定义的工作效率分析图表功能。（分析配置中自定义配置）

2.3.9 *敏感词功能*

敏感词分词：用户配置敏感词分析规则中需要分析哪些内容。目前支持正则身份证和手机号码，或者匹配数据集。

敏感词规则：用户配置需要记录敏感词的数据源以及规则。

敏感词明细：提供已命中的敏感词明细查询。

敏感词分析：提供自定义的敏感词分析图表功能。（分析配置中自定义配置）

2.3.10 *图表功能优化*

4.2 的报表功能只支持明细记录查询以及一级维度的聚合图表。为适应客户对各类数据更细粒度的查询需求，将报表的图表功能进行拆分，可以支持更多、更复杂的数据分析聚合。

指标：指标主要用于聚合字段划分数据后，获得的需要统计的结果。在 4.2 中，默认指标是分类后的数量。如按用户名聚合，出来的数据即每个用户名的记录条数。现版本支持数量、求和、求最大值、求最小值。

图表：图表用于配置分析的字段，并配置数据指标的方式，为用户提供分析结果。目前图表支持柱状图、饼图、折线图、列表。可以支持多指标对比分析，支持指标结果过滤。图表可以自定义条件。

报表：报表功能取消了原来直接计算图表功能，采用引用图表的形式。可以在一个报表中引入多个分析图表，方便客户对某一类问题进行综合分析

仪表盘：提供综合图表展示功能，可以自定义增加图表到仪表盘，仪表盘可在首页切换显示，主要用于某个场景的数据分析展示。

2.3.11 *自定义分析功能*

在 4.2 中有客户反馈我们提供的分析图表并不是需要的。所以设计将分析图表独立，进行一个个场景分析。

分析类型：提供自定义分析类型的设定

分析配置：分析功能依赖图表，主要通过引用图表功能提供数据分析。一个分析里面可以支持多个图表，并且搜索条件可以自定义。

目前支持 4 个模块的分析配置：用户画像，风险分析，工作效率分析，敏感词分析。

3. 下个版本计划

已知的列入到下个版本的功能：

网页内容敏感词分析

linux 桌面录屏

server 端提供 rsyslog 转发及 splunk 转发

邮件附件上报分析

打印队列行为监控

ES 服务器监控

图表、分析功能优化

4.5 版本计划未定稿，仅供参考。