ALEIYE

让 | 大 | 数 | 据 | 更 | 简 | 单

Aleiye 大数据分析平台操作手册

北京数介科技有限公司

2016-5

目录

第1章 概述

1.1 编写目的

该文档的目的是描述 aleiye 产品的用户使用说明，其主要内容包括：

- 产品介绍
- 运行环境
- 产品功能与操作说明

1.2 编写范围

该文档定义了 aleiye 产品的使用说明,主要描述了 aleiye 产品的功能介绍、所用的搜索命令、配置说明以及产品操作使用流程。

1.3 名词解释

数据类型 ：对导入平台中的数据进行类型划分，类型名称可自定义，在搜索中，以数据类型为基础进行结果查询。

表格数据：是指有每行数据字段间都是用固定的分隔符进行分割的数据。典型的代表是 csv 文件。

文本采集：通过采集器采集数据。

数据上传：将数据以文件形式一次性导入平台。

1.4 参考文档

《Aleiye 产品 V3.2 产品安装手册》

《Aleiye 产品 V3.2 检索命令手册》

第2章 产品介绍

2.1 什么是 Aleiye

Aleiye 是一款强大的数据引擎，能够实时收集、索引、统计所有 IT 设备（服务器、网络设备、应用程序、数据库）和基础结构（物理、虚拟和云中）生成的日志数据，包括完成数据分析、日志分析以及业务数据分析，并通过图形化的方式展现出来。

除了能满足一般日志分析软件功能外，Aleiye 还支持海量信息搜索及更多的功能，比如分散式搜寻、计划告警、权限控制等。

Aleiye 使用机器学习和数据挖掘等技术，对跨越多个系统的复杂事件进行关联分析，监视、分析和预警设备运维状态，避免服务性能降低或中断，从而降低运维成本，减少运维风险，提升服务质量。

2.2 Aleiye 功能简介

整合数据：Aleiye 可根据 IT 数据格式来整合运维数据、安全信息及数据、应用程序数据。日志可能分散在各个主机或集群中随机的数据源中，Aleiye 可以从一台机器或一组机器中快速搜索出所有的日志。

搜索日志：Aleiye 强大的搜索功能可以从繁杂的日志数据里查出任何您想要的内容。

提取关键信息：Aleiye 可以自定义正则表达式，从每一行中提取变量，并对其进行过滤、分组和聚合的功能。

可视化：Aleiye 可以通过命令行和函数为用户提供可视化图表，包括饼图、柱状图、行列图等图表。

告警提醒：Aleiye 可以对事件告警进行设置，通过阀值、数据异常来触发邮件提醒。

预警：Aleiye 可以通过不同的预警算法帮助用户预测单值或多值字段可能会发生的值。形成未来趋势的报告，使用户提前了解并做出相应的措施。

数据挖掘：Aleiye 可以从大量数据中挖掘隐含的、未知的并具有潜在价值的信息。帮助用户做出正确的决策分析。

特点：

Aleiye 是一种高扩充性且通用的数据引擎，具有以下特点：快速查找日志信息、数据关联分析、支持各种平台和系统。

2.3 Aleiye 使用对象和应用领域

2.3.1 使用对象

Aleiye 是一款多功能的搜索引擎，用途广泛，适合以下不同类型的用户，Aleiye 可提供更便捷、更全面的服务。

基础方面：系统管理员、网络工程师、软件开发工程师等。此类用户可使用 Aleiye 检查服务器问题，了解其配置及监测用户活动。还可将搜索结果转变成告警，警报服务器性能阈值、关键性系统错误及负载。

运营方面：服务台及应用支持人员、产品经理等。此类用户可根据 Aleiye 事件分析状态，维护或验证分析产品模型，为产品定位调整作为依据条件。

决策方面：部门经理、副总裁、CIO(首席信息官)等，此类用户可使用 Aleiye 构建报告和仪表板来检测并汇总其 IT 基础构建和业务的健康状况、性能、活动及容量。

2.3.2 应用领域

Aleiye 为各行业提供专业的智慧型管理解决方。不仅可以应用于传统行业如医疗、教育、金融服务、零售等，而且还被 IT 行业所使用，如远程通信、电信、金融、信息安全等领域。

Aleiye 聚焦于金融行业的运营管理、销售支持和商业模式创新。

帮组企业建成统一的数据平台，实现“针对正确的人，在正确的时间，正确的方式，提供正确的信息”的目标，形成企业智慧型管理模式。

第3章 功能列表

模块	功能	描述
首页	数据展示	展示事件总数、今日处理事件、保存周期等相关处理数据，其中今日处理量涉及 lincense 权限。根据不同 lincense 所展示的数据则不同
	功能展示	在首页中，展示出搜索、仪表盘、上传方式、数据类型、等功能的快速入口。
	首页仪表盘	可以在首页，添加已创建好的仪表盘。
创建数据类型	正则表达式	通过正则表达式的方式对数据源进行解析工作，更适用于非机构化数据。
	切分符	通过固定切分符如“逗号”“空格”“分号”等对数据进行解析工作，更适用于结构化数据。
选择上传方式	数据上传	通过文件上传方式，将数据一次性导入平台中。
	Syslog 上传	通过 syslog 协议将数据导入平台中。
	文本采集	通过采集器，对数据进行实时采集。
	SNMP 协议采集	通过 SNMP 协议将数据导入平台中。
	数据库连接	可以与数据库进行对接，将数据库中的数据导入平台中，目前支持 orcale、mysql、sqlsever 和 db2 三种数据库类型。
告警设置	计划告警	对搜索结果，以某一时间点，周期性的对其进行监控，超过阀值便会触发告警动作。
	实时告警	基于数据源中的规则触发告警动作。
搜索	原语搜索	通过原语语句，进行数据搜索查询，搜索结果具有实时性。
	SQL 搜索	基于 SQL 语句进行搜索，可通过离线任务对大数据量进行搜索。支持 95 标准。
数据监控	采集器监控	以时间单位，对采集器采集的事件数量进行监控。

	文本采集	对所采集路径的事件数进行监控，并可以对其进行删除等操作。
	Syslog 采集监控	监控 syslog 协议开启状态，及其采集事件数，并可以对 syslog 协议进行开启关闭等操作。
	SNMP 采集监控	监控 SNMP 协议开启状态，及其采集事件数，并可以对 SNMP 协议进行开启关闭等操作。
	数据库连接监控	监控数据库采集数据采集数据状态，及其采集的事件数，并可以手动同步数据库中数据。
仪表盘	仪表盘	可以关联报表中的数据，将多张报表在一个仪表盘中进行展示。
字典库	字典库	可以创建字典表，当前支持文件上传和页面自定义两种方式将字典表导入平台中，文件上传支持 CSV 格式。
数据关联	数据关联	基于数据类型，通过字段与 key 值相关联，从而实现字典中的值补全到数据中。
数据挖掘	数据挖掘	基于数据类型和检索语句可以进行算法展现，目前支持曲线预测、直线预测、异常点分析和异常模式关联四种。

第4章 用户操作说明

4.1 首页

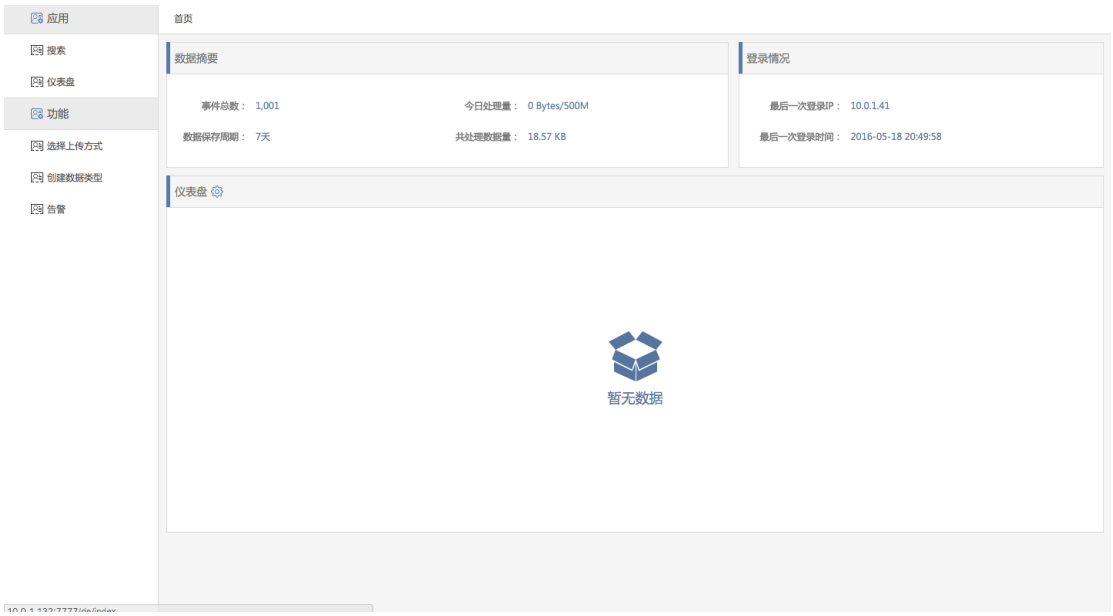


图 4-1

在首页中，我们为用户提供数据可视区域，在该区域中，用户可以添加某一仪表盘内容在首页进行展示，其次，我们将平台中的主要应用和功能的入口在首页进行展示，方便用户快速进入不同功能点。

- 数据摘要：数据摘要所统计的事件数均为实时数据，其中，今日处理量与产品授权相关联，每天的数据处理量不能超过上限值。
- 首页仪表盘：可以将仪表盘关联到首页，并在首页进行仪表盘的设置功能。

4.2 创建数据类型

图 4-2

添加数据的过程中，需要先创建数据类型，再选择上传方式，将创建好的数据类型关联到上传方式中，从而完成整个添加数据的流程。创建数据类型最主要的功能就是对数据的解析，在该版本中，我们对数据类型的切分分为两种，一种为正则表达式，一种为切分符切分。

4.2.1 数据类型管理

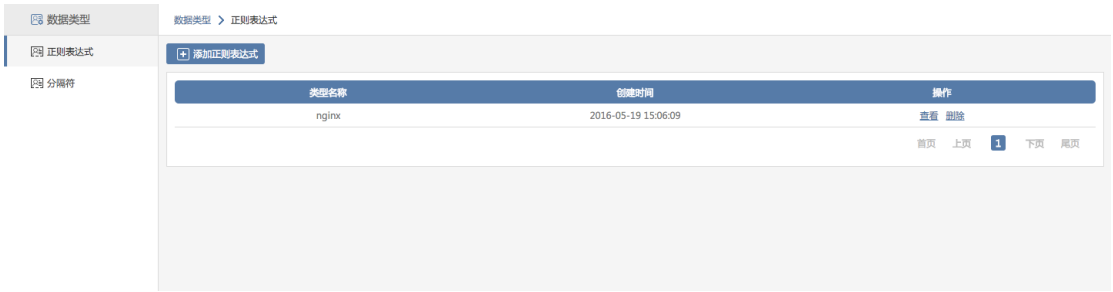


图 4-3

对创建好的数据类型，我们需要对其进行相应的管理，主要对数据类型进行增删改查等操作。

添加数据类型：在管理页面，可以添加新的数据类型，点击添加按钮，弹出数据类型配置页面，在 4.2.2 和 4.2.3 两个章节，会对配置参数详细描述。

删除数据类型：创建好的数据类型，在没有被引用的情况下，是可以进行删除操作，如果一旦被上传方式被引用的话，则不能执行删除操作的。

编辑数据类型：编辑功能与删除功能规则一致，在数据类型被引用的情况下

是不可以进行编辑，如果该数据类型没有被引用，则可以进行编辑操作，进入编辑页面，可以对创建的配置项进行修改。

4.2.2 正则表达式切分

图 4-4

在创建数据类型配置页面中，可以通过正则表达式的方式对数据解析，其中配置参数包括：文件上传、名称、正则表达式、解析预览。

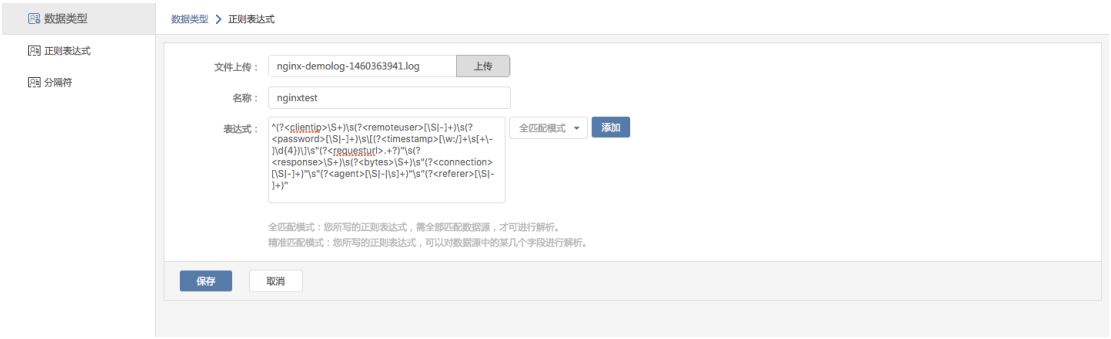


图 4-5

文件上传：在做解析的时候，需要用户上传样例数据，样例数据大小限制在 5MB 以内，暂不支持压缩格式，只支持文本格式的文件。

名称：为数据类型自定义名称，当前名称不支持中文，且不能重名。

正则表达式：正则表达式分为两种模式，一种为全匹配模式，一种为精准拼配模式，每种模式下的正则表达式可以添加多个，系统会按照正则表达式的顺序逐一进行解析，当多个正则表达式中的字段相同的话，后面的字段会覆盖前面的字段。

基于全匹配模式下的正则表达式，解析过程中数据会与正则表达式进行完全

匹配，一旦正则表达式与数据无法匹配，则整个解析过程失败。

基于全匹配模式，填写的正则表达式必须与数据完全匹配才可以进行解析，如果正则表达式中有一个字段是无法解析的，则所有数据是无法解析成功的。其解析的原始数据：123.116.170.191 - - [11/Apr/2016:16:25:07 +0800] "GET /de/login;jsessionid=32669AB3DC720A5DD440486A88A52030 HTTP/1.1" 200 4230 "-" "JoeDog/1.00 [en] (X11; I; Siege 2.78)" "-"，正则表达式为：^(?<clientip>\S+)\s(?<remoteuser>[\S|-]+)\s(?<password>[\S|-]+)\s\[(?<timestamp>[\w:/]+\s[+\-]\d{4})\]\s"(?<requesturl>.+?)"\s(?<response>\S+)\s(?<bytes>\S+)\s"(?<connection>[\S|-]+)"\s"(?<agent>[\S|-|\s]+)"\s"(?<referer>[\S|-]+)"。该正则表达式可以完全对数据源进行解析，如图 4-6。

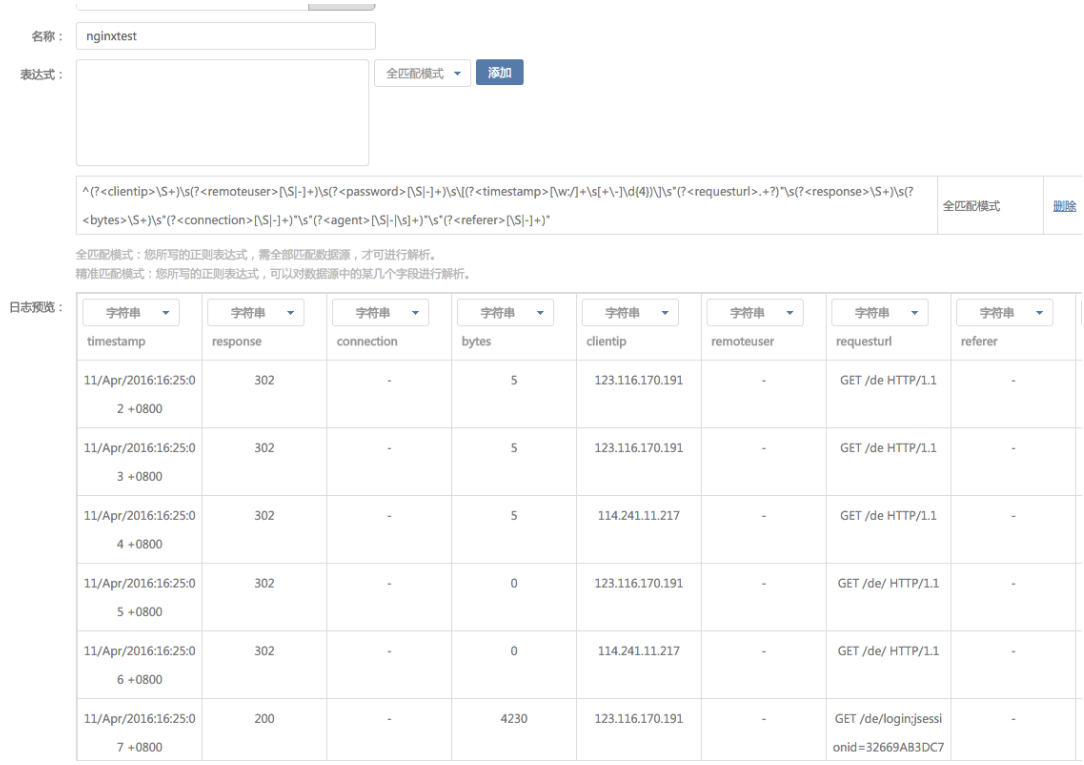


图 4-6

基于精准匹配模式下的正则表达式，所填写的正则表达式，无需全部匹配数据源，只要能够解析出一个字段，该解析流程即为成功，其解析的数据源：10.10.255.1 <189>1001: apr 19 13:52:56.366: %LINK-3-UPDOWN: Interface GigabitEthernet1/1, changed state to up，正则表达式为：(?<status>\sd[\w]+|\su[\w]+)，该正则表达式只能解析出上述数据源中的 up 信息，所以基于精准匹配模式，

可以做到基于字段进行逐一解析。如图 4-7。

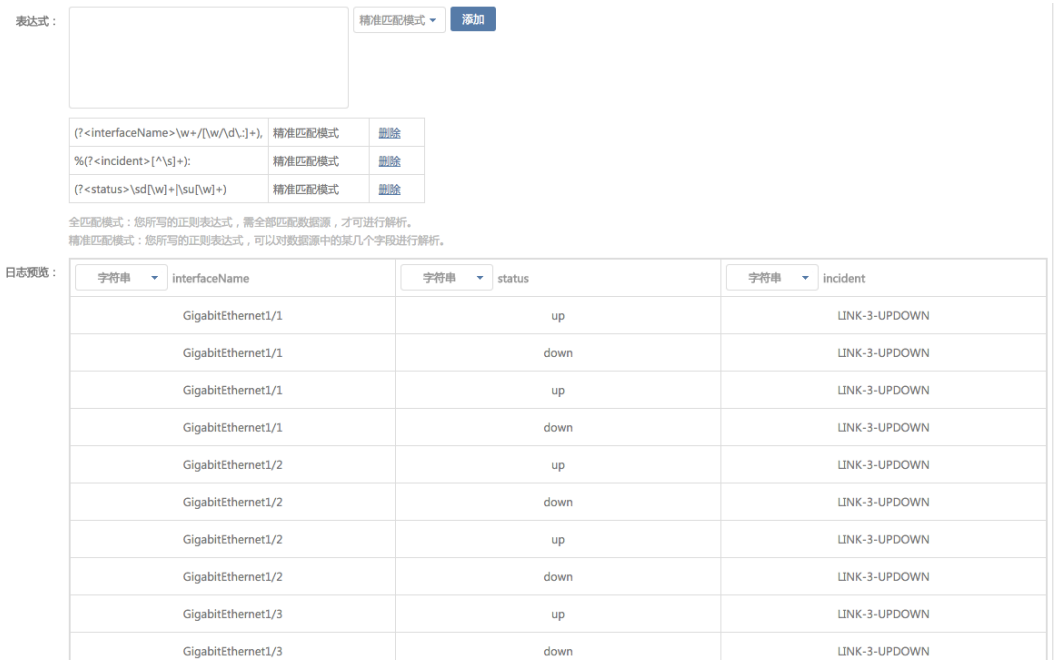


图 4-7

解析预览：解析预览会针对上面的配置项，将解析结果进行展示，预览结果只展示上传样例数据中前 10 条数据。在预览中，可以配置解析后字段的类型，当前支持三种字段类型：字符串、数字、日期、键值对。

系统会基于日志中的业务时间和日志入库的时间来判断先后顺序，所以需要在选择日期类型的字段是否是业务时间，如果不勾选，则按照入库时间对数据进行排序。

业务时间为日志中的时间，即数据所产生的时间，入库时间为数据进入平台的时间。

系统会提供多种日期格式适配数据中的日期，最终转换成“2016-07-13 19:23:20.18”，并提供不同时区选择，同时也可以自定义日期格式和所在时区。如图：4-8。

图 4-8

4.2.3 分隔符切分

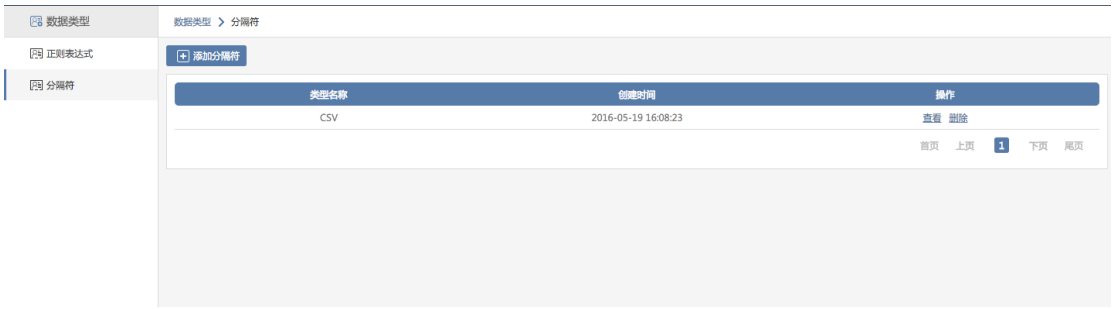


图 4-9

在创建数据类型的过程中，系统支持切分符的方式对数据进行解析，其中配置参数包括：文件上传、名称、解析预览。

文件上传：与正则表达式中的文件上传功能一致，支持 5MB 以下的数据，且支持文本格式的文件，不支持压缩的数据格式。

名称：创建数据类型的唯一标识，不支持中文，且不能重名。

解析预览：数据上传后，会看到该数据的预览状况，系统会默认读取上传的

数据中前 10 条记录，并给该数据定义一个默认字段“field0”。基于默认字段，可以进行出拆分和合并操作，将默认字段拆分成多个字段，并自定义字段名称和字段类型。如图：4-10。

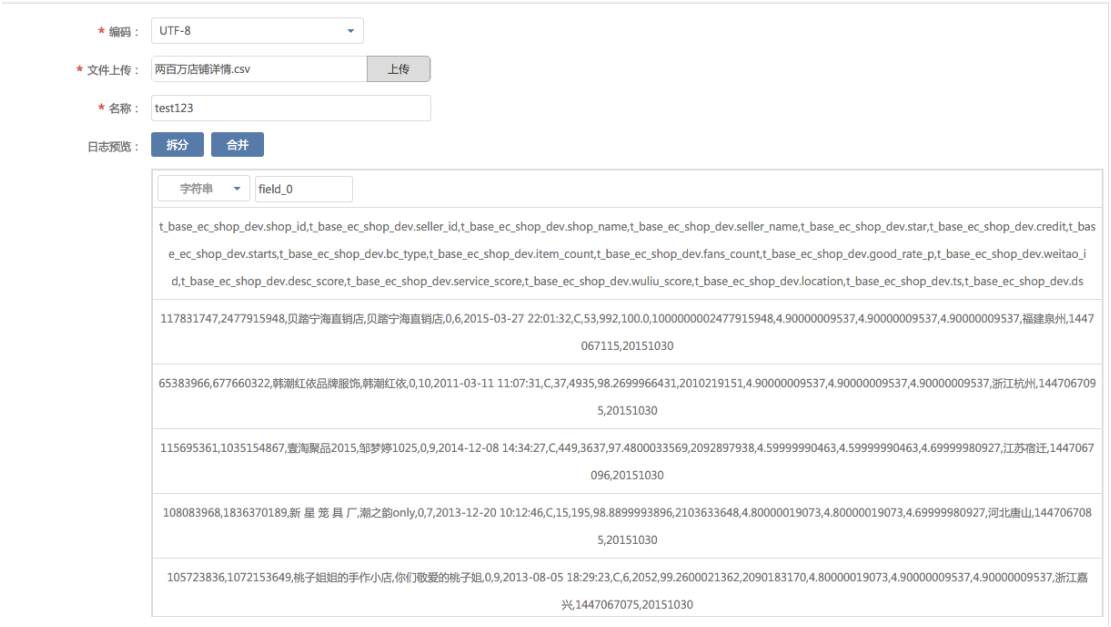


图 4-10

通过拆分功能，可以通过固定的分隔符对预览中的数据进行结构化拆分，4-10 中的数据样例，我们可以通过“，”对其进行结构化拆分，点击拆分按钮，如图 4-11，可以自定分隔符，并确认需要切分的字段名称（field0），确认之后，系统会按照选中的分隔符对选中的字段进行切分操作。

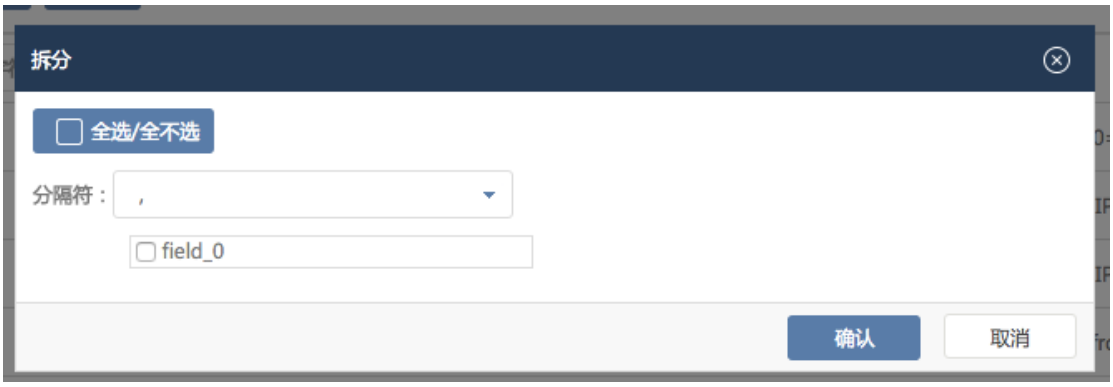


图 4-11

拆分后结果如图 4-12，拆分后，还可以对字段进行合并的操作。

字符串 field_0	字符串 field_1	字符串 field_2	字符串 field_3	字符串 field_4	字符串 field_5	字符串 field_6
t_base_ec_shop_dev.shop_id	t_base_ec_shop_dev.seller_id	t_base_ec_shop_dev.shop_name	t_base_ec_shop_dev.seller_name	t_base_ec_shop_dev.star	t_base_ec_shop_dev.credit	t_base_ec_sh
117831747	2477915948	贝踏宁海直销店	贝踏宁海直销店	0	6	2015-03-2
65383966	677660322	韩潮红依品牌服饰	韩潮红依	0	10	2011-03-1
115695361	1035154867	壹淘聚品2015	邹梦婷1025	0	9	2014-12-0
108083968	1836370189	新星笼具厂	潮之韵only	0	7	2013-12-2
105723836	1072153649	桃子姐姐的手作小店	你们敬爱的桃子姐	0	9	2013-08-0
100611846	444824331	小李子出品	金融叉叉	0	7	2012-10-1
64252791	652646031	茂鸿铁艺家饰工艺厂	小芳阁88	0	10	2010-12-3
59908599	283371501	两个宝贝童品店	lkx5425	0	12	2009-11-2

图 4-12

点击合并按钮，选中需要合并的字段，并通过分隔符对字段进行合并如图 4-13。

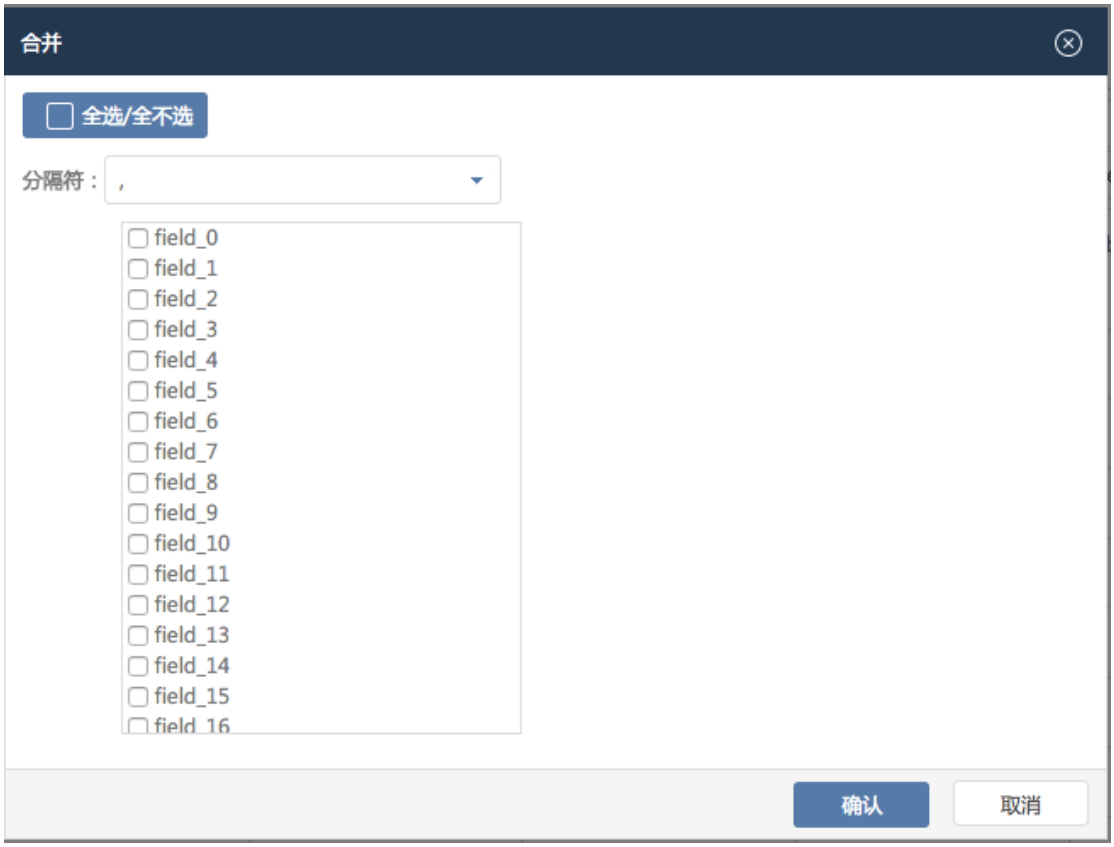
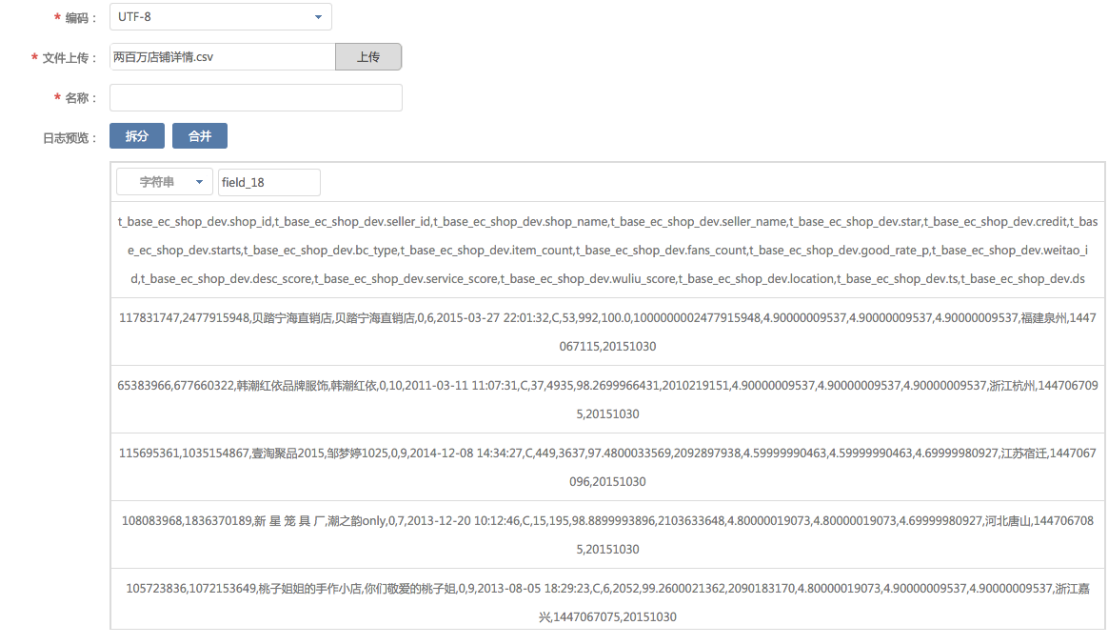


图 4-13

将刚解析后的结果，可以通过合并的功能，还原回一个字段，选中所有字段，

选择“,”的分隔符，点击确认后，所有字段会合并成一个字段，如图 4-14.



4.3 选择上传方式

当前版本中，提供五种上传方式，包括文本采集、数据上传、syslog 采集、SNMP 采集和数据连接五种上传方式，用户根据实际业务需求选择不同的上传方式，其中，文本采集和 syslog 采集、SNMP 采集和数据连接都需要下载 Aleiye 所提供的采集器，并在要采集的设备上进行安装部署，数据上传方式无需采集器，只需要将数据上传到平台中即可。

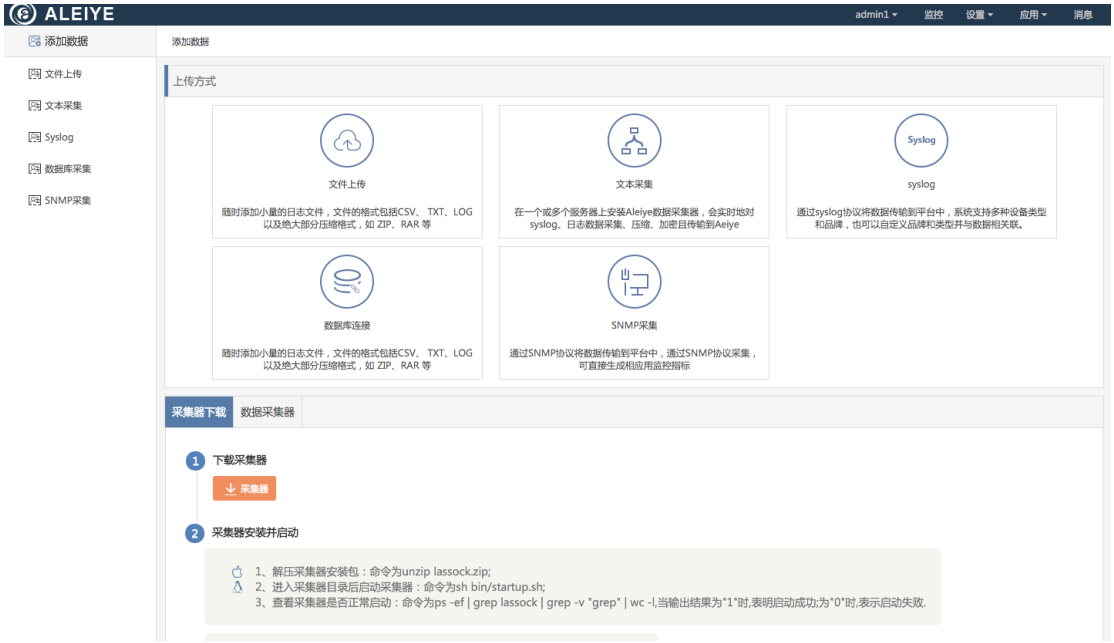


图 4-12

4.3.1 文件上传

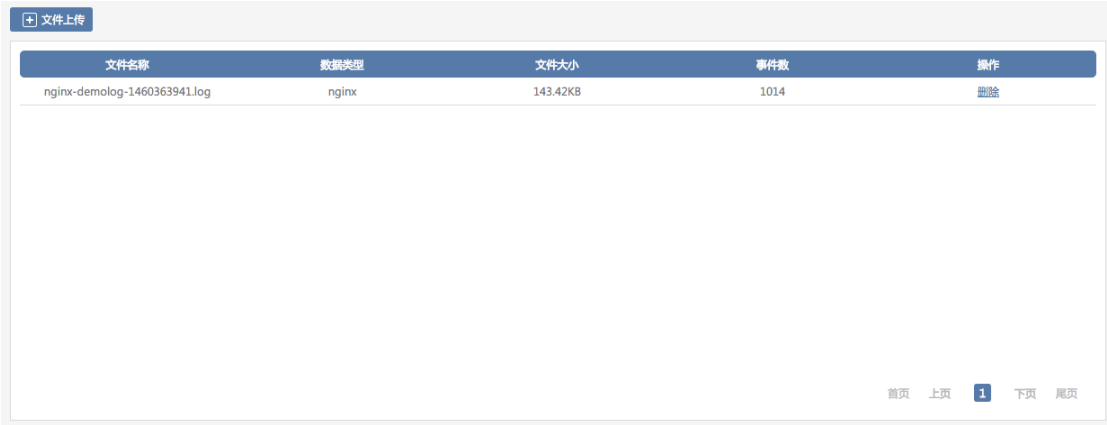


图 4-13

通过文件上传讲述少量的数据导入到平台中，上传文件最大不超过 5MB，且支持文本格式的文件。

文件上传管理页面：文件上传列表页，主要是对已创建的上传方式进行增删改查的功能。

删除：可对已创建的数据上传记录进行删除操作。

文件上传配置页面：通过添加按钮入口跳转到文件上传的配置页面，配置页面包括数据类型、选择文件两个配置项。

数据类型：在下拉菜单中，不仅包含系统自带的数据类型，还包含自定义的数据类型，如果已有的数据类型不能满足需求，可以通过“创建数据类型”按钮跳转到创建数据类型的页面，进行自定义数据类型。

选择文件：文件上传的方式，可以点击上传文件的输入框，通过浏览目录的方式选择上传的文件，也可以同过拖拽的方式将文件拖拽到页面中，进行上传。文件上传支持多个文件、文件夹（文件夹中可以有多个文件夹，且可不限级别）和一级压缩文件(压缩文件文件中不能再有压缩文件)。当多个文件或文件夹时，拖拽框中应该展示出每个文件或文件夹的名称。总上传文件的大小不能超过 5MB。

图 4-14

4.3.2 文本采集

文本采集是通过Aleiye自主研发出的采集器对设备中的日志进行实时采集，在使用文本采集的前提需要安装部署采集器，在文本采集页面中下载采集器，如图 4-15。

图 4-15

下载采集器后，需要将采集器安装部署到所需要采集数据的设备上，根据不同系统，安装步骤不同。

Liunx、mac 系统安装步骤：

1、解压采集器安装包：命令为 unzip lassock.zip;

2、进入采集器目录后启动采集器：命令为 sh bin/startup.sh;

3、查看采集器是否正常启动：命令为 ps -ef | grep lassock | grep -v "grep" | wc -l,当输出结果为"1"时,表明启动成功;为"0"时,表示启动失败。

Windows 系统安装步骤：

1、采集器安装包解压：利用解压工具进行解压操作;

2、启动采集器：进入 lassock 目录,运行 bin 目录下的 startup.cmd。

3、查看采集器是否正常启动：在采集器目录－logs 文件夹下查看采集器启动日志。

通过上述步骤，安装采集器完成后，系统中监控也米娜可以查看安装好的采集器状态，如图：4-16。

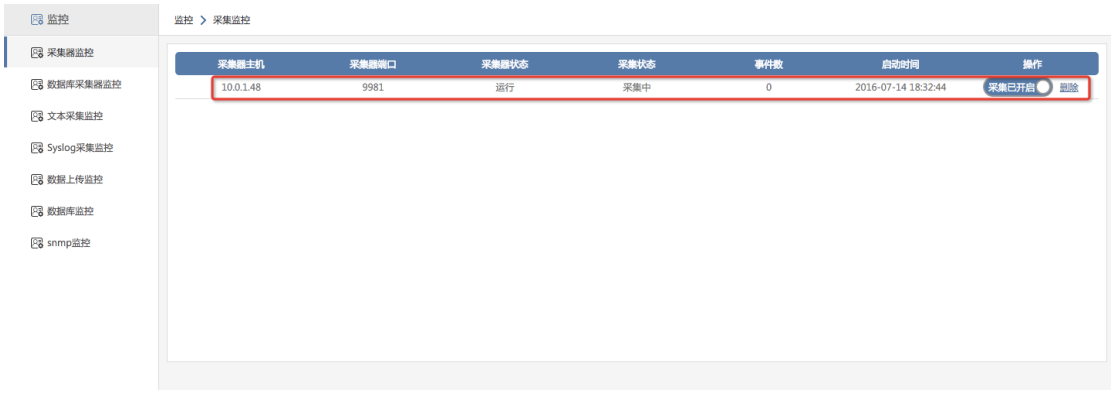


图 4-16

如果出现上图记录，则代表采集器安装部署成功，安装好的采集器默认为开启状态，采集器状态为运行且采集状态为采集器中，则证明，采集器运转正常，可以通过文本采集采集的方式上传数据了。

在上传方式中，选中文本采集，进入文本采集方式配置页面，如图 4-17。

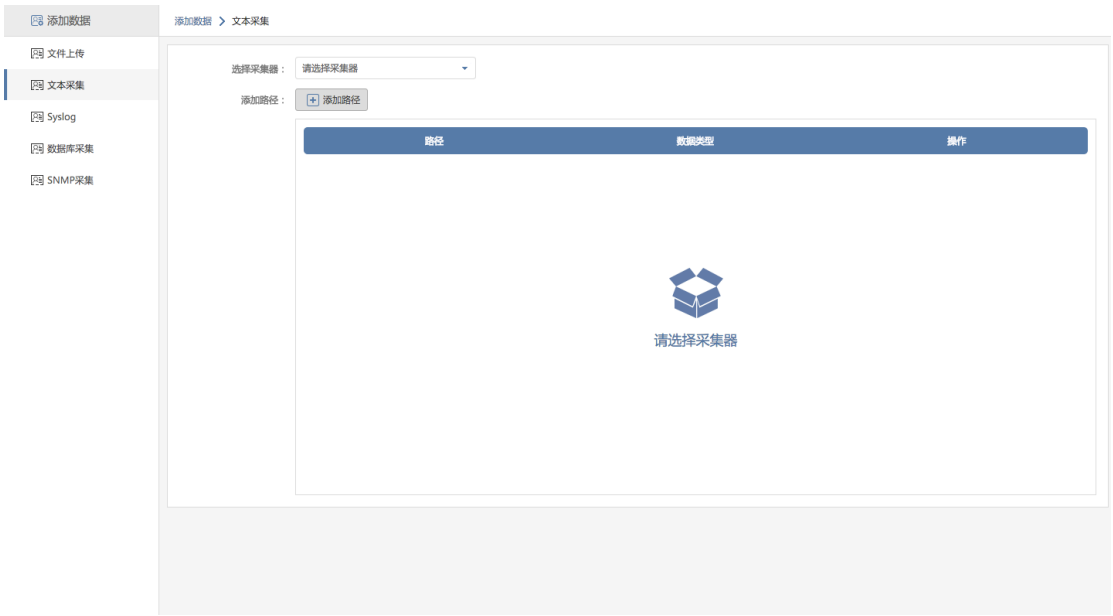


图 4-17

在做配置之前，需要选择采集器，在安装好采集器后，选中之前安装好的采集器后，便可以在该采集器下，配置数据路径，该采集器会根据路径实时采集路径下的数据。

配置路径的过程中，需要选择数据类型，即通过正则表达式或者分隔符方式对数据进行解析后的数据类型，同时，还需要选择编码，当前支持 ASCII、GBK、GB2312 四种编码，最后，录入需要采集器路径并保存配置，采集器会根据配置实时采集数据，如图：4-18。

图 4-18

配置好文本采集后，可以通过监控页面中的采集器监控进行查看，如果采集

事件数量出现变换则证明，数据已经开始进入平台，并可以进行搜索功能的使用。如图 4-19。

图 4-19

4.3.3 syslog 采集

syslog 常被称为系统日志或系统记录，是一种用来在互联网协议（TCP/IP）的网络中传递记录信息的标准。Aleiye 通常采用 syslog 协议的方式采集网络设备、安全设备以及服务器设备的系统日志数据；

- 启动 syslog 采集流程

通过 syslog 协议采集数据，同样需要先下载采集器（采集器下载安装详见 4.3.1 章节）

- 启动 syslog 采集

采集器安装部署完成后，需要在 syslog 监控页面手动开启 syslog 协议。Syslog 协议默认通过 UDP 协议 514 端口进行接收；如果数据源端未采用默认的协议以及端口时，用户需要通过［syslog 采集监控］模块进行修改；

- syslog 配置

选择采集器：由于 syslog 需要基于采集器进行采集，所以在创建或者查看 syslog 采集记录，需要选中采集器，在该采集器下可以创建 syslog 采集或查看。如图 4-20。

图 4-21

添加 syslog：添加 syslog 采集需要配置数据类型、编码、数据源 IP、设备厂商和设备类型。如图 4-21

图 4-21

数据类型：syslog 数据采集后需要通过数据类型进行数据的解析以及管理，定义新的数据类型需要通过［创建数据类型］模块进行添加；

编码：用户需要选择数据的编码以保证日志数据中的中文字符能够正确解析；Aleiye 平台对 UTF-8、ASCII、GBK、GB2312 等常用编码均可支持；用户也可自定义编码格式；

数据源 IP 地址：用户添加 syslog 时需要定义数据源的 IP 地址，用户也可一次配置多个数据源 IP 地址，多个 IP 地址之间用“;”间隔；

设备厂商和设备类型：添加 syslog 时需要配置设备厂商以及设备类型，用于数据分析时进行日志信息的标记；用户可通过设备厂商管理页面，创建新的关联；

如上述流程所示，syslog 采集流程配置完成；用户可通过数据检索、数据可视化等模块进行数据的分析以及展示；

为确保数据的完整性，Aleiye 采用避免数据丢失的机制进行数据采集；当用户在［syslog 监控］模块启动 syslog 采集后，syslog 数据就开始采集，采集到日志数据系统默认以数据类型“syslog”进行解析和管理；当数据解析失败后则通过数据类型“err”进行管理；

“syslog”默认配置：编码：UTF-8

“syslog”默认解析字段：

A_hostName:采集器所在的服务器地址；

A_processTime:系统接收数据的时间，也称入库时间；

A_remoteIp:数据源 IP 地址；

A_source:数据类型，Aleiye 管理数据的基本单位；

A_timestamp:业务时间，用来记录日志数据产生的时间；

Facility：特性，由 2 个或 2 个以上大写字母组成的代码，用来表示硬件设备、协议或系统软件的型号。

FacilityCode:特性编号，取值范围为 0～23；每个编号分别代表设备的不同模块；

Severity：严重性，范围为 0～7 的数字编码，表示了事件的严重程度。

SeverityCode:严重性编号，每个消息优先级有一个十进制的严重级别编号。取值范围为 0～7；

4.3.4 SNMP 采集

基于SNMP协议进行数据采集,在通过SNMP采集数据之前,需要先部署Aleiye提供的采集器，启动采集器后，需要在 SNMP 监控页面手动启动 SNMP 协议，开启协议需要配置端口号。(开启 SNMP 协议详见 SNMP 采集监控说明)。

- 采集器下载并启动

基于 syslog 协议，对数据进行采集，同文本采集一样，在采集数据前，需要先下载 Aleiye 所提供的采集器。(采集器下载并以及启动的方法请参见页面提示)

- 启动 SNMP 采集

采集器安装部署完成后，需要在 SNMP 监控页面手动开启 SNMP 协议(具体操作见 SNMP 采集监控功能)。SNMP 协议默认通过 UDP 协议 161 端口进行接收；如果数据源端未采用默认的协议以及端口时，用户需要通过［SNMP 采集监控］模块进行修改；

- SNMP 配置

选择采集器：在创建 SNMP 协议采集前，需要选择采集器所在的 IP 地址，采集器部署的设备上，必须启动 SNMP 协议，基于该设备上的采集器进行 SNMP 协议采集。如图 4-22。

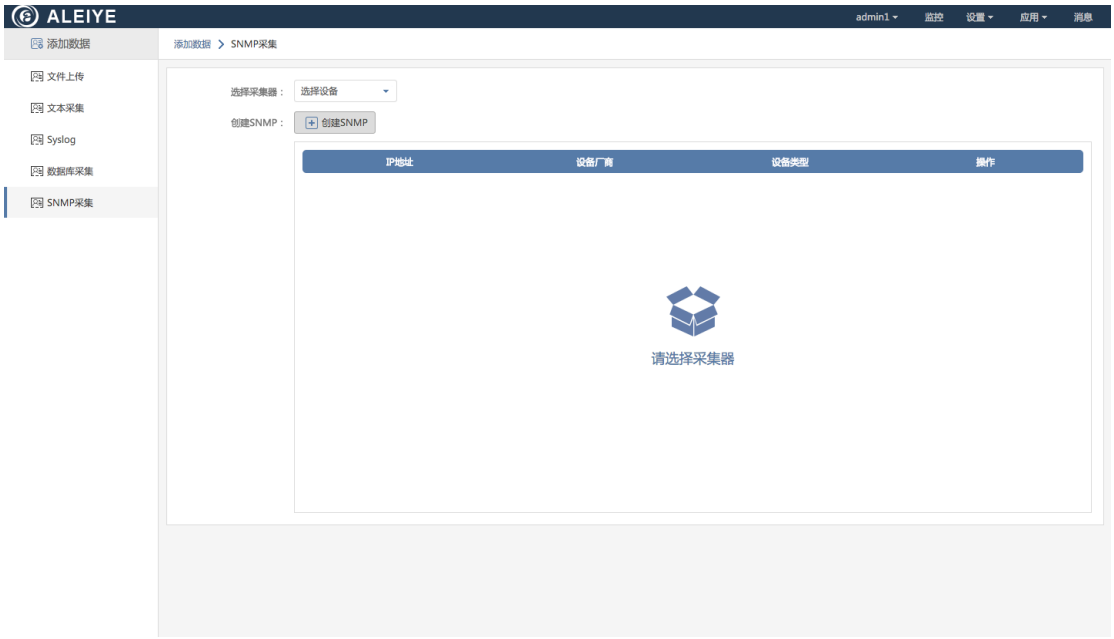


图 4-22

SNMP 协议采集配置：配置 SNMP 采集时，系统需要了解你所采集的 SNMP 的 IP 地址，配置完设备信息后，还需要配置 SNMP 协议版本，当前版本支持 V2 和 V3，当设备信息和 SNMP 协议配置完成后，可以选择所获取的指标信息，其中包括：设备状态（cpu 使用率、内存使用率、设备温度）、端口信息（设备 IP、设备名称、设备型号、端口名称、端口描述、端口 IP、端口配置状态、MTU）、端口统计信息（半双工以太网利用率、全双工以太网利用率、端口发送率、接收丢包率、发送丢包率、接收错误率、发送错误率、接收广播/多点发送包速率、发送广播/多点发送包速率）、端口状态信息（配置状态、当前状态）。如图 4-23

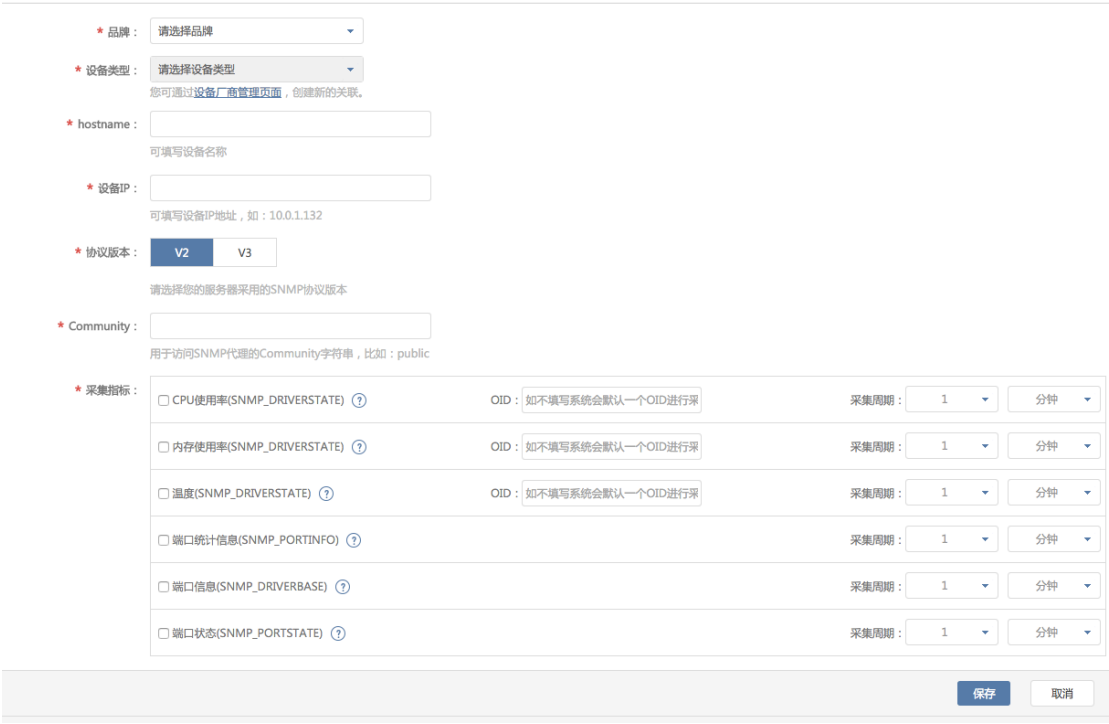


图 4-23

配置完成后请点击［确认］进行页面保存；

如上述流程所示，SNMP 采集流程配置完成；用户可通过数据检索、数据可视化等模块进行数据的分析以及展示；

SNMP 采集各指标的数据系统以默认的数据类型进行解析和管理；具体指标对应的数据类型如下：

SNMP 采集指标	数据类型	字段名称及含义
CPU 使用率	a_SNMP_DRIVERSTATE	driverIp（设备 IP） driverName（设备名称） cpu（cpu 平均使用率） memory（内存平均使用率） temperature（设备温度）
内存使用率		
温度		
端口统计信息	a_SNMP_PORTINFO	driverIp（设备 IP）

		driverName（设备名称） portIp（端口 IP） portName（端口名称） sysuptime（监控时间） halfDuplexEthernet（半双工以太网利用率） fullDuplexEthernet（全双工以太网利用率） portSend（端口发送率） portReceive（端口接收率） reveivePacketLoss（接受丢包率） sendPacketLoss（发送丢包率） reveiveError（接收错误率） sendError（发送错误率） reveiveBroadcast（接收广播/多点发送包速率） sendBroadcast（发送广播/多点发送包速率） inSecSpeed（接收到的每秒速率） periodFlow（接收到的每秒总和） inflow（当前接收到的流量） inLastFlow（上次接收到的流量） outSecSpeed（发出的每秒速率） outPeriodFlow（发出到的每秒总和）

		outflow（当前发出到的流量） outLastFlow（上次发出到的流量）
端口信息	a_SNMP_DRIVERBASE	driverIp（设备 IP） driverName（设备名称） portIp（端口 IP） portName（端口名称） portDes（端口描述） conState（端口配置状态） portMtu（端口 MTU）
端口状态	a_SNMP_PORTSTATE	portIp（端口 IP） portName（端口名称） conState（端口配置状态） curState（端口当前状态）

4.3.5 数据库连接

Aleiye 支持四种数据库类型的连接，包括：oracle、mySQL、SQLsever 和 db2。在通过与数据库创建连接后，可以支持多库多表增量或全量的方式将数据库中的数据导入到平台中。在创建数据库连接前，需要安装部署 Aleiye 提供的数据库采集器，通过数据库连接、参数配置，实现数据库数据的采集。

在选择上传方式页面中，点击数据库采集器下载按钮下载。如图 4-22

图 4-22

下载后，进行采集器安装部署操作，针对不同操作系统，安装部署步骤也不同。

Liunx、mac 系统安装步骤：

1、解压采集器安装包：命令为 unzip dbcollect.zip;

2、进入采集器目录后启动采集器：命令为 sh bin/startup.sh;

3、查看采集器是否正常启动：命令为 ps -ef | grep dbcollect | grep -v "grep" | wc -l,当输出结果为"1"时,表明启动成功;为"0"时,表示启动失败.

Windows 系统安装步骤：

1、采集器安装包解压：利用解压工具进行解压操作;

2、启动采集器：进入 dbcollect 目录,运行 bin 目录下的 startup.cmd。

3、查看采集器是否正常启动：在采集器目录－logs 文件夹下查看采集器启动日志;

数据库连接管理页面：在数据库管理页面，当选中数据库采集器后，可以对

该采集器下的所有数据库连接记录进行增删改查的操作，当数据库导入数据过程中，其状态为任务中，在该状态的情况下，是不可以对该条记录进行删除和编辑操作。

在添加数据库连接时，需要先选择安装好的数据库采集器，所添加的配置项都将在该数据库采集器下进行执行。如图 4-23。

图 4-23

数据库配置共分为三部分；数据库连接、自定义表、增量全量导入配置。如需要将某一数据库中的所有表全部导入到平台中，需要先建立访问该数据库的权限，如图 4-24

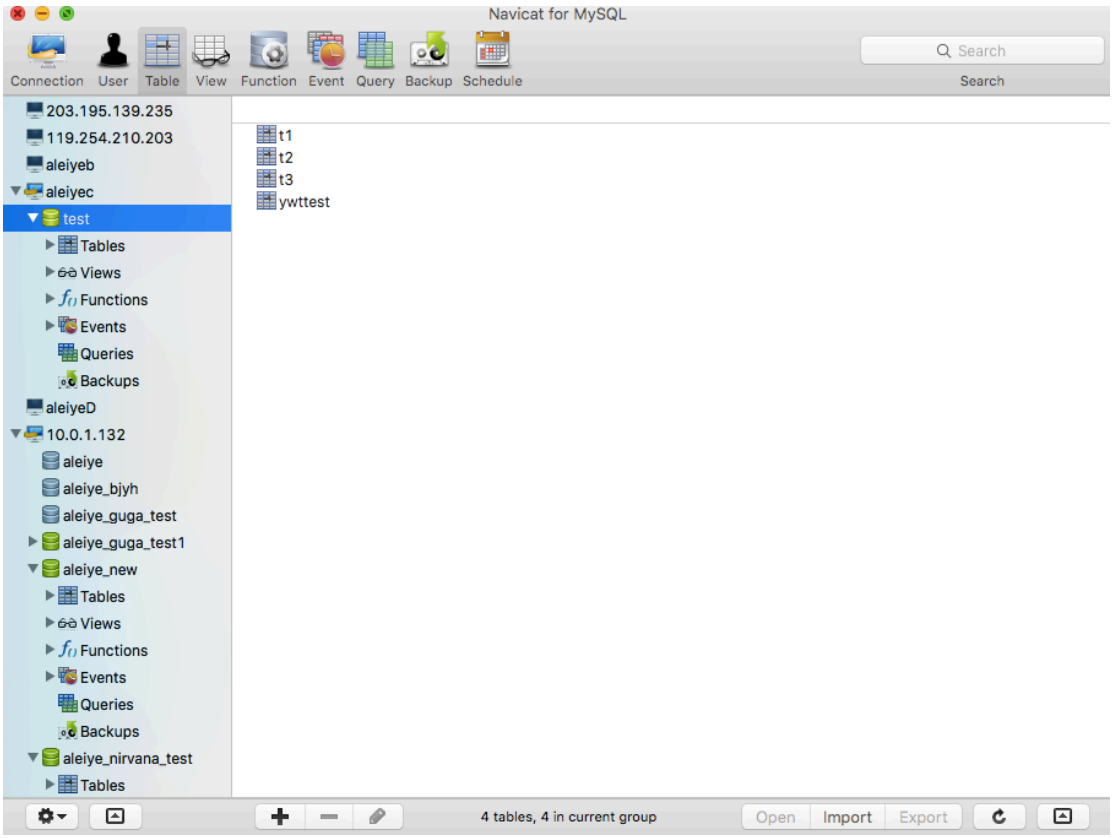


图 4-24

- 数据库连接配置：

URL 数据库别名：导入的数据库进行唯一标识，数据库 URL 地址别名命名规则为字母开头且支持字母、数字和下划线，不支持中文，长度不得超过 6 个字符（不可重复）。

数据库名称：需要填写正确的数据库名称，如上图中的“test”。

数据库类型：当前版本支持四种数据库类型，分别是 ORCALE、MYSQL、SQLSEVER 和 DB2。样例中的数据库为 SQLSEVER 的类型。

数据库 URL 连接：需要访问该数据库是所需要的数据库 URL 地址，例如：jdbc:mysql://10.0.1.132:3306/test?useUnicode=true&characterEncoding=utf-8。

用户名密码：需要录入相关数据库的用户名和密码。

schema 信息：当选择 ROCALE 数据类型时，需要选择 schema 信息，一般情况下，schema 信息为大写的数据库用户名。

上述配置完成后，需要点击验证，验证过程基于上述的配置参数尝试与该数据库进行连接，如果连接失败，需要重新配置该参数直到与数据库连接成功后，才可以进行入库的配置项，如图 4-25。

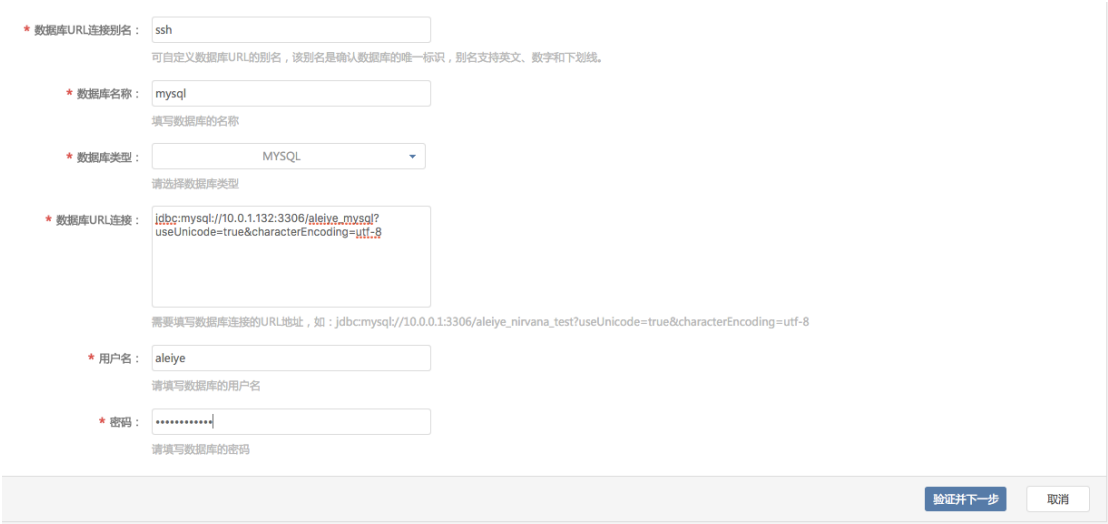


图 4-25

- 自定义表导入：

数据库表：通过数据库连接的配置，系统会读取该库中的所有表信息，并进行展示，你可以自定义需要导入哪些表。

自增字段：在选择增量导入或全量导入的情况下，需要选择自增字段，自增字段一般为数字类型，如果不是数字类型则不可以选择为自增字段，系统会基于自增字段进行数据采集。

全量导入：如果选中全量导入的话，系统会将选中的数据库表一次性全部导入到平台中，在全量导入选项中，可以选择性的选择自增字段，如果选择自增字段的话，会相应提高数据库导入的性能。

增量导入：增量导入时根据自增字段的值，周期性的采集数据库中的数据，所以在增量导入中，必须选择自增字段。如图 4-26

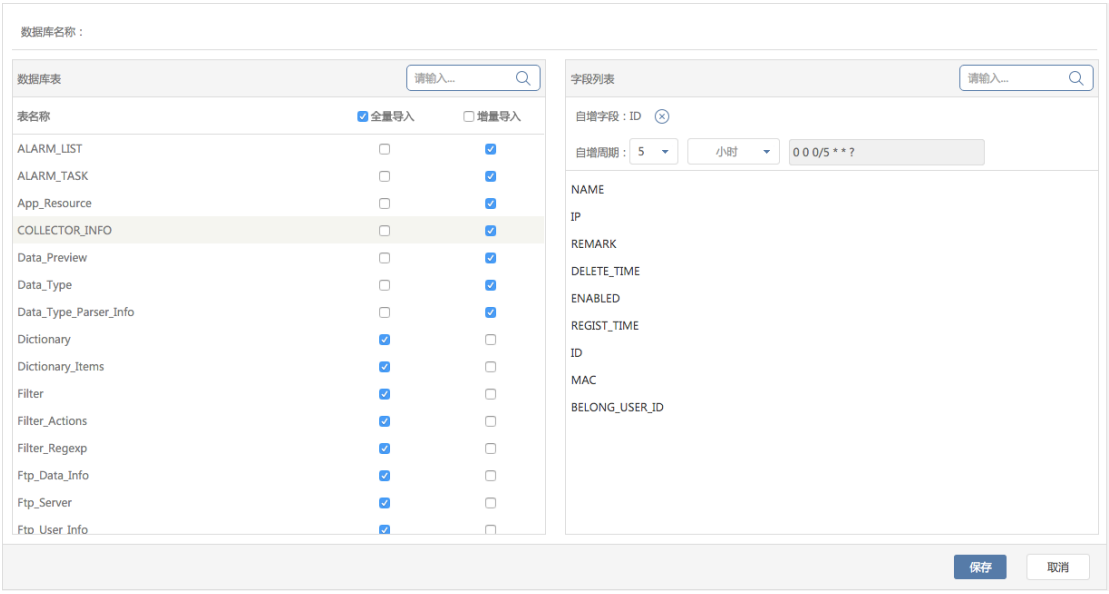


图 4-26

当数据库导入到平台后，表的信息即为数据类型，可以通过数据类型在搜索中进行查询。在原有表名前，增加了 A_db 的前缀，保证表的唯一性，同时，数据库中的字段钱准为数据库别名、表明、字段名。

4.4 监控

Aleiye 数据分析平台可以在数据采集过程中进行采集数据量和状态等监控功能，会基于采集器和采集方式两个维度进行监控，如图 4-27

图 4-27

4.4.1 采集器监控

针对不同的上传方式，aleiye 提供了两种采集器，一种为数据采集器，针对文本文件、syslog 和 SNMP 三种上传方式应用，另一种则为数据库采集器，主要是与数据库进行对接采集数据，对两种采集器，需要进行启停等操作，并实时监控其采集的数据量。

- 数据采集器监控：

在 4.3.2 章节中，已经对数据采集器安装部署进行了详细说明，采集器在安装部署完成之后，在监控页面可以查看到该采集器的记录，由于页面存在延迟，可以通过页面中刷新按钮刷新页面，从而在页面获取安装部署好的采集器信息。

采集器主机：监控采集器所部署安装设备的 IP 地址，可以清楚的看到采集器所在位置。

采集器端口：采集器端口默认为 9981。

采集器状态：采集器状态共分为两种，运行和未运行，在采集器安装部署完成成后，该状态默认为运行，如果采集器状态为未运行的话，说明后台没有启动采集器，需要重新安装部署采集器，在未运行状态下，无法对采集器进行启停的操作。

采集状态：采集状态与采集器状态相关联，如果采集器状态是未运行的话，则采集状态就是未采集，一旦采采集器状态是运行，则采集器状态为采集中。

事件数：当采集器开始采集器数据后，会实时的监控采集数据的事件数。

启动时间：可以查看采集器启动的时间点。

操作：可以对采集器进行启停和删除两种操作，当采集器状态是未运行时，是不可以进行启停操作的；执行删除功能后，会删除采集器的进程，所以需要重新安装部署才可以再次启动采集器。

图 4-28

- 数据库采集器：

在 4.3.5 章节中，对数据库采集器的安装部署进行了详细的说明，采集器安装部署完成后，可以对采集器进行监控，当前版本，数据库采集器还不支持手动启停和删除功能，在之后版本会逐步支持。

采集器主机：同数据采集器一样，该项监控采集器所部署在设备的 IP 地址，清楚的知道采集器所在的位置。

采集器状态：该项与数据采集器状态一样，共分为两种，运行和未运行，在安装部署完成后，数据库采集器的状态默认为运行状态。如果数据库 采集器所在设备重启或者关机的情况下，采集状态是为运行，需要重新安装部署采集器。

事件数：当数据库采集器开始采集数据的过程中，可以实时监控采集的事件数。

启动时间：可以查看采集器启动的时间点。

图 4-29

4.4.2 采集器方式监控

- 文本采集监控：

文本采集监控是对文本采集方式对其进行监控，以采集器为维度，监控每个采集器下的所有路径的采集状况。

路径：在创建文本采集方式的时候，需要添加路径，采集器会基于该路径采集数据。

路径状态：路径状态共分为两种，正常和异常，如果出现异常状态的话，说明添加的路径存在问题，系统没有找到相应的路径，需要在文本采集方式中查看或修改路径。

接收数：对采集器采集到数据进行监控。

发送数：对采集器向平台发送的数据进行监控，当发送数和接收数之间有较大的差值的话，说明采集器获取到数据，但是没有发送到平台，产生数据堆积问题。

最后上报时间：平台接收最后接收到数据的时间。

图 4-30

- Syslog 采集监控：

Syslog 采集监控是对 syslog 采集方式进行监控的，syslog 协议是基于数据采集器，如果平台没有安装采集器的话，sysylog 协议是没有的，在 4.3.3 章节中也提到过，在使用 syslog 协议传输数据时，必须安装采集器。

采集器 IP：由于 syslog 协议是基于采集器，所以该项展示出所安装部署的采集器地址，以便对该采集器开启 syslog 协议。

Syslog 状态：syslog 状态共分为两种，启动和停止，默认 syslog 为停止状态。

端口：syslog 协议的端口号，默认状态下为 514 端口，如果需要修改可通过编辑功能对其进行修改。

协议：syslog 协议默认为 UDP，当前支持 UDP 和 TCP 两种协议。

事件数：通过 syslog 协议传输数据的事件数量监控。

最后上报时间：平台最后接收到数据的时间点。

操作：通过操作功能，可以手动启停 syslog 协议，并可以通过编辑按钮，修改端口号和协议。

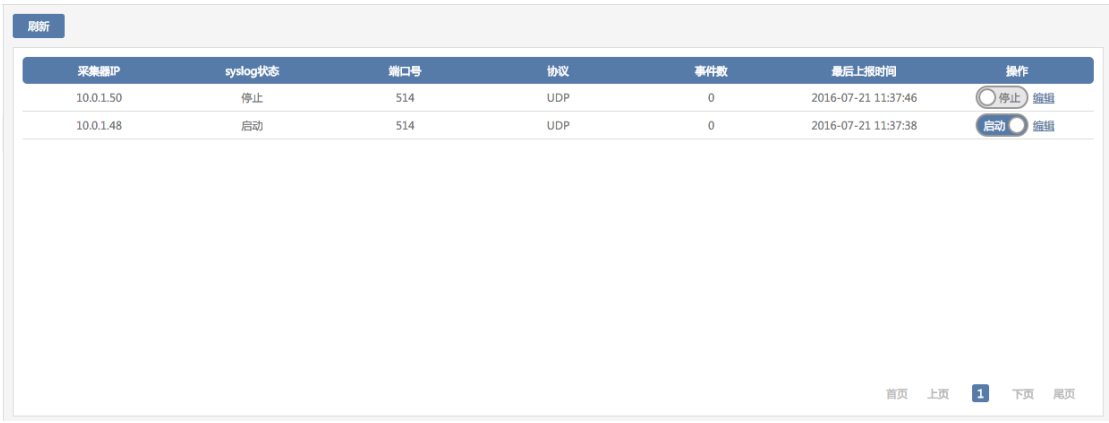


图 4-31

- 文件上传监控

文件上传监控室对文件上传方式进行监控，通过文件上传的方式将数据导入到平台后，可以通过该监控功能查看上传的文件名称大小等。

文件名称：通过文件上传方式上传的文件名称。

数据类型：该上传的文件使用的是哪种数据类型进行的解析。

文件大小：上传文件的大小。

事件数：上传的文件事件数据量。

操作：可以将该条记录进行删除，该删除不能删除真正的数据，而是将记录进行删除。

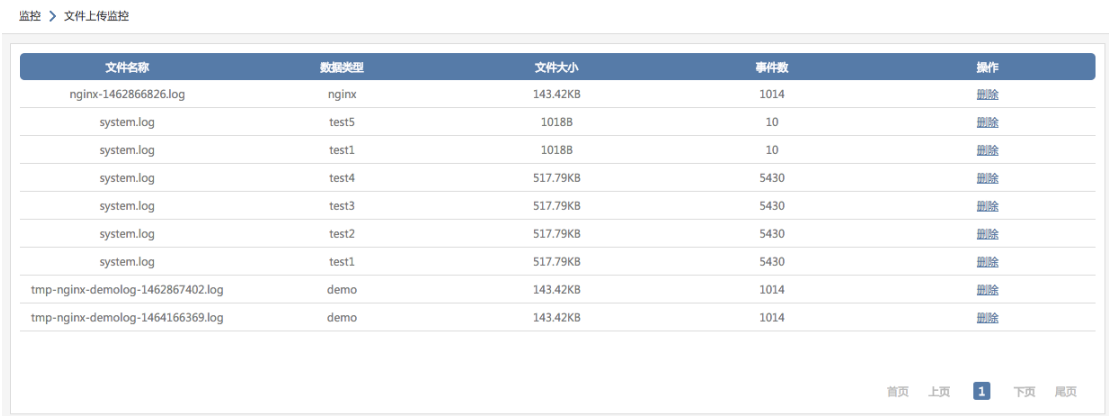


图 4-32

- SNMP 采集监控：

SNMP 采集监控是对 SNMP 采集方式进行监控的，SNMP 协议是基于数据采集器，如果平台没有安装采集器的话，SNMP 协议是没有的，在 4.3.4 章节中也提到过，在使用 SNMP 协议传输数据时，必须安装采集器。

采集器 IP：由于 SNMP 协议是基于采集器，所以该项展示出所安装部署的采集器地址，以便对该采集器开启 SNMP 协议。

SNMP 状态：SNMP 状态共分为两种，启动和停止，默认 SNMP 为停止状态。

端口号：SNMP 默认端口号为 161，不可修改。

协议：SNMP 协议默认为 UDP，不可修改。

事件数：通过 SNMP 协议传输数据的事件数量监控。

操作：可以通过操作功能，可以手动启停 SNMP 协议。

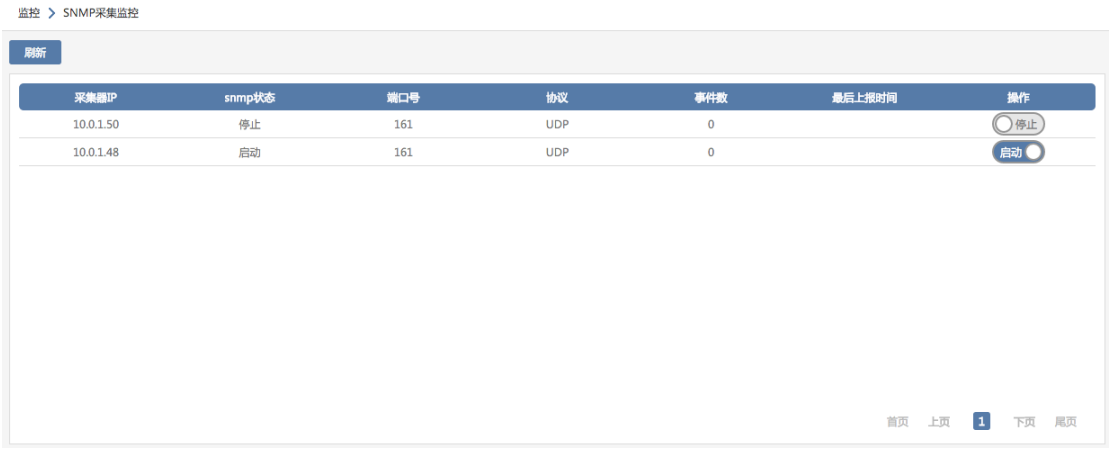


图 4-33

- 数据库采集监控

数据库采集监控，以安装部署好的数据库采集器为维度，监控每个采集器下的所有数据库连接记录。

采集器 IP：记录数据库采集器所安装部署设备的 IP 地址，方便快速找到采集器所在的位置。

数据库类型：在 4.3.5 章节中，已经提到，共支持的数据库类型包括 MYSQL、SQLsever、orcale 和 DB2 四种。

状态：数据库采集状态共分为两种，采集中和采集完成，当状态为采集中的

情况下，是不可以进行删除操作的。

事件数：对数据库采集数据的事件数量进行监控。

图 4-34

4.5 原语搜索

Aleiye 的搜索功能可以将 Aleiye 整合的不同源、不同类型的数据进行检索，以图表或报告的形式展示给用户。检索是通过输入关键字结合时间范围的方式，以满足不同业务对数据检索的需求，原语搜索只能搜索仅 7 天或 30 天的数据，从保证数据的实时性。检索结果可以导出为 CSV 格式。

搜索页面，按 Aleiye 的搜索语言，在搜索栏中输入搜索信息（搜索命令详见如何搜索），按回车键或点击搜索栏右端的搜索图标启动搜索。点击搜索栏右侧按钮打开下拉菜单，可选择预设或者自定义时间范围进行搜索（注：数据文件如果有时间戳，时间范围以时间戳为准，否则以入库时间为准）。如图 4-25。

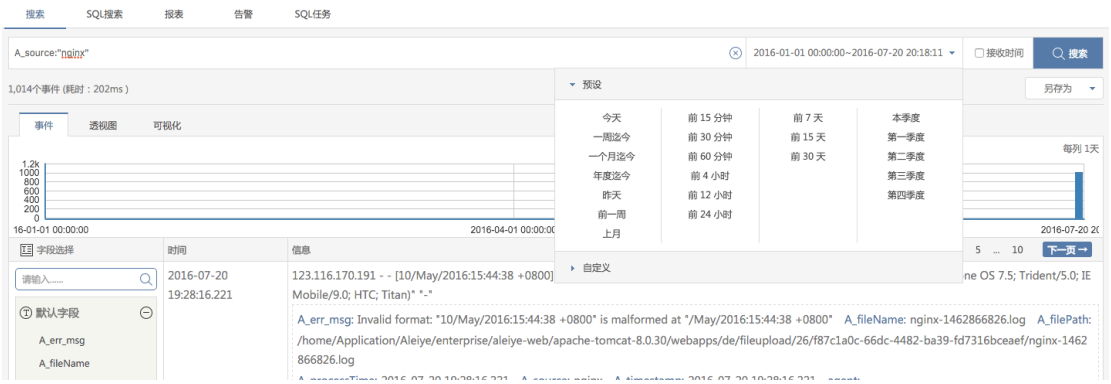


图 4-35

点击搜索栏打开辅助搜索界面，辅助搜索提供保存的搜索条件、保存的报表，可帮助用户快速搜索，如图 4-36。

保存的搜索条件：搜索条件可以进行保存，并在该功能中进行快捷操作。

保存的报表：显示通过搜索结果保存的报表，点击报表可重现报表结果。

数据类型名称：在搜索前，必须选择一个数据类型进行搜索功能，当创建完数据类型并上传到平台后，便可以通过该模块选择创建好的数据类型进行搜索操作。

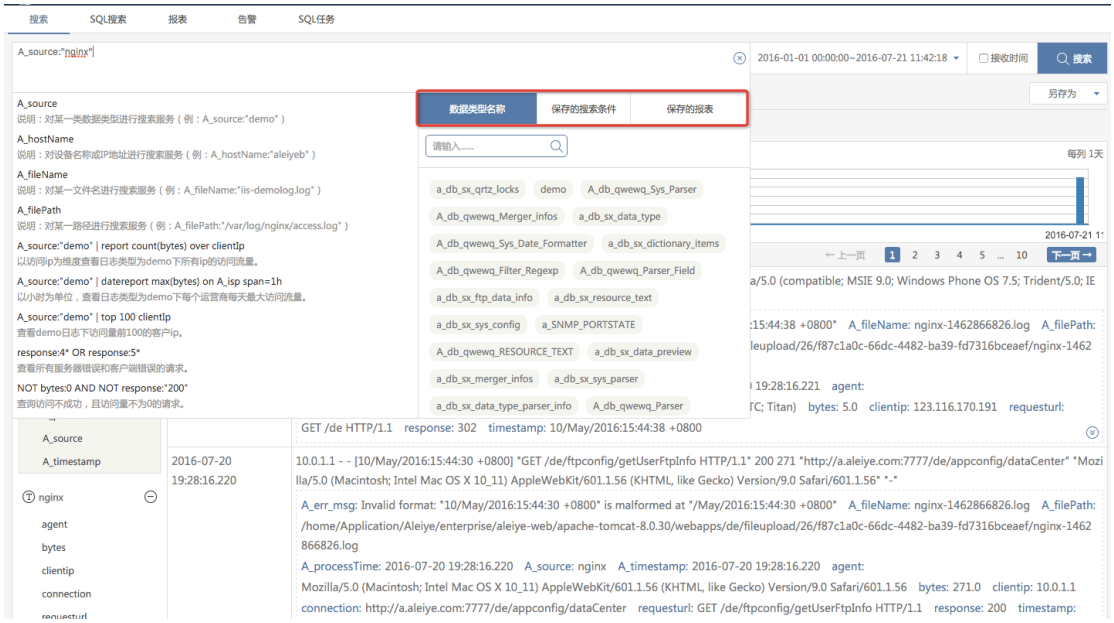


图 4-36

在搜索框中，输入语句，如：A_source:”nginx”的命令，即搜索之前所创建的 nginx 数据类型，基于搜索语句，返回结果就是 nginx 数据类型下的所有解析后的数据如图 4-37。

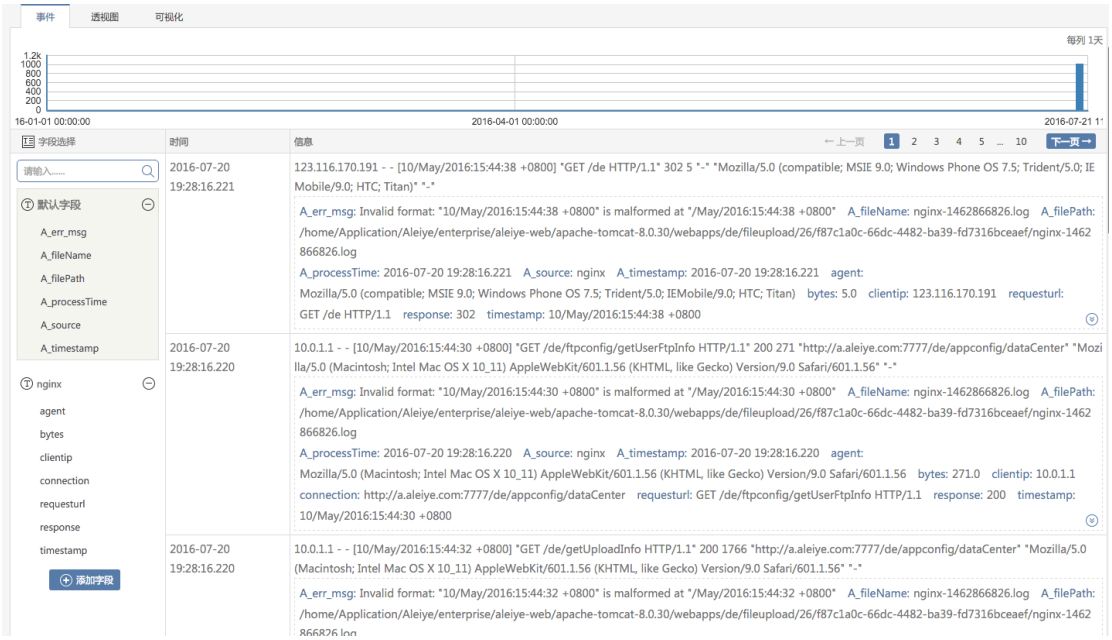


图 4-37

在事件信息列表中，字段选择列分为默认字段和自定义字段，该展示结果基于搜索语句结果展现，其中默认字段为 Aleiye 自带字段，不管任何数据类型都会包含该字段，自定义字段则为搜索的数据类型所包含的字段。默认字段包含：

A_filePath ：采集路径 通过采集端采集文件有值对应 例如：/usr/aaa/bbb.log

A_fileName ：上传文件此字段有值 例如：my.log

A_source ：数据类型

A_hostname ：只有启用采集端 此字段有意义

A_processTime：数据接收时间戳，在数据没有业务时间的话，会以数据入库的时间定义时间戳。

A_timestamp：数据中的业务时间戳。

A_err_msg：错误信息，如果数据在进入平台的过程中，出现例如解析错误等会议该字段进行统计展示。

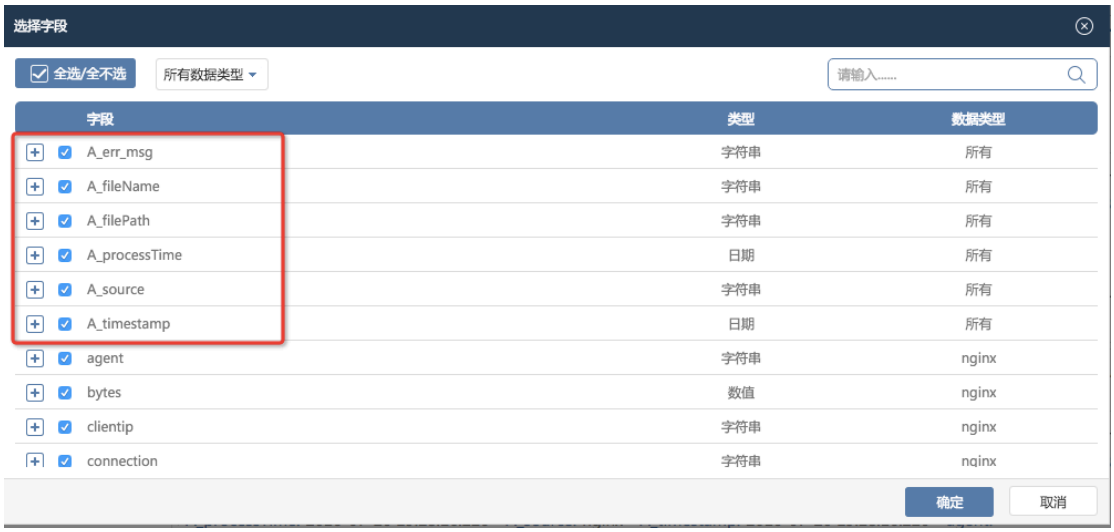


图 4-38

除默认字段外，还包括基于搜索结果后的字段，如搜索“A_source:”nginx””后，除了上述默认 A_开头的字段外，还有 nginx 解析后的字段，如 agent、bytes、clientip 等。如图 4-39。

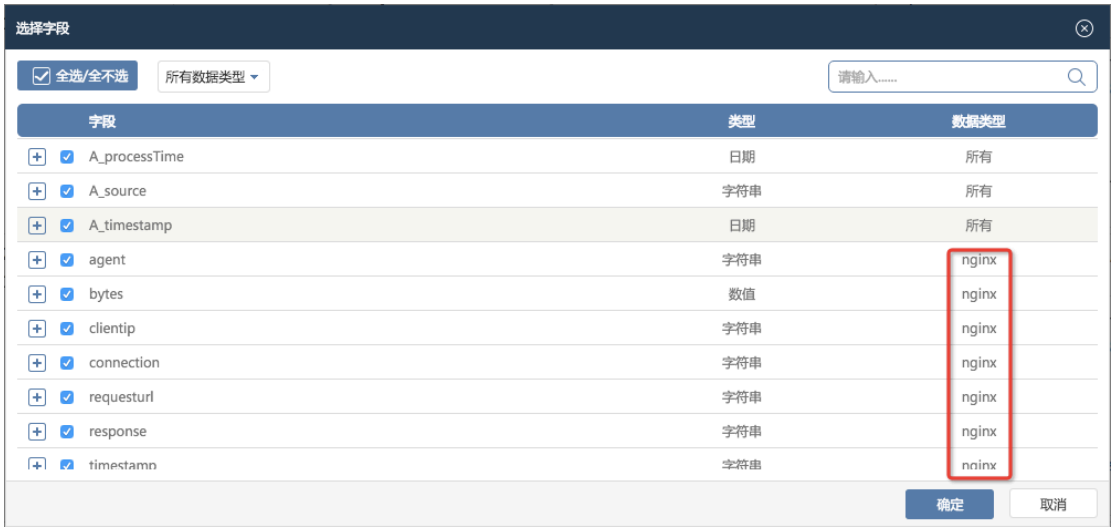


图 4-39

在信息列表中，点击信息栏下方的下拉图标显示事件信息详情。在事件信息列表中点击字段名，Aleiye 可根据字段名过滤事件信息。如图 4-40。

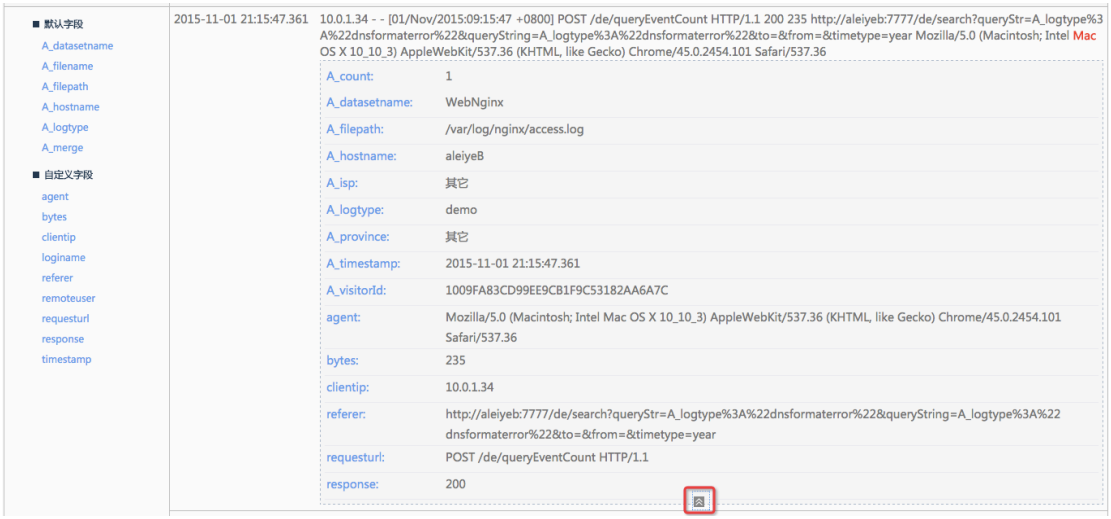


图 4-40

点击“透视图”图标可切换到透视图操作页面，根据搜索结果，将结果中的字段进行展现，并通过拖拽形式，快速形成图表。透视图标签只有在报表命令、EVAL 语句和 SQL 语句中才会触发，如果只是条件搜索的话，该功能无法使用，如图：4-41。

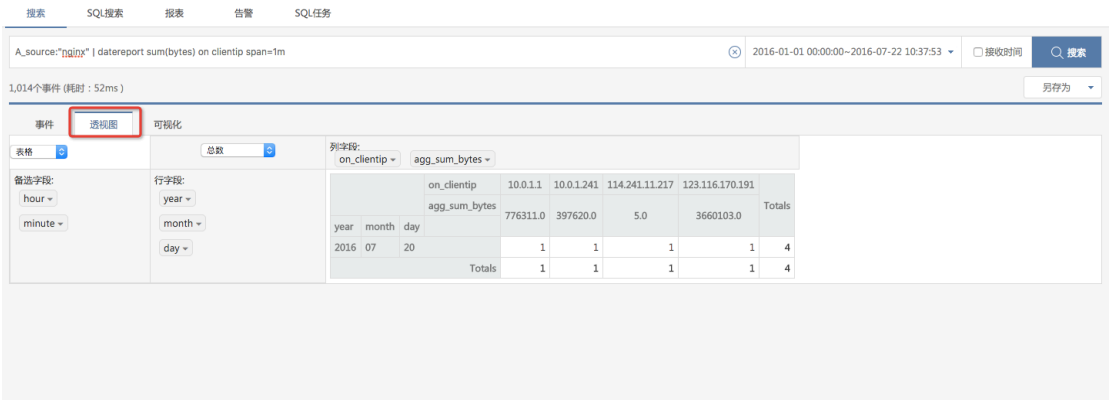


图 4-41

点击“数据可视化”图标可切换至数据可视化页面，将数据的各个属性值以多维数据的形式展示，只有在报表命令的情况下才会触发该功能。如图 4-42

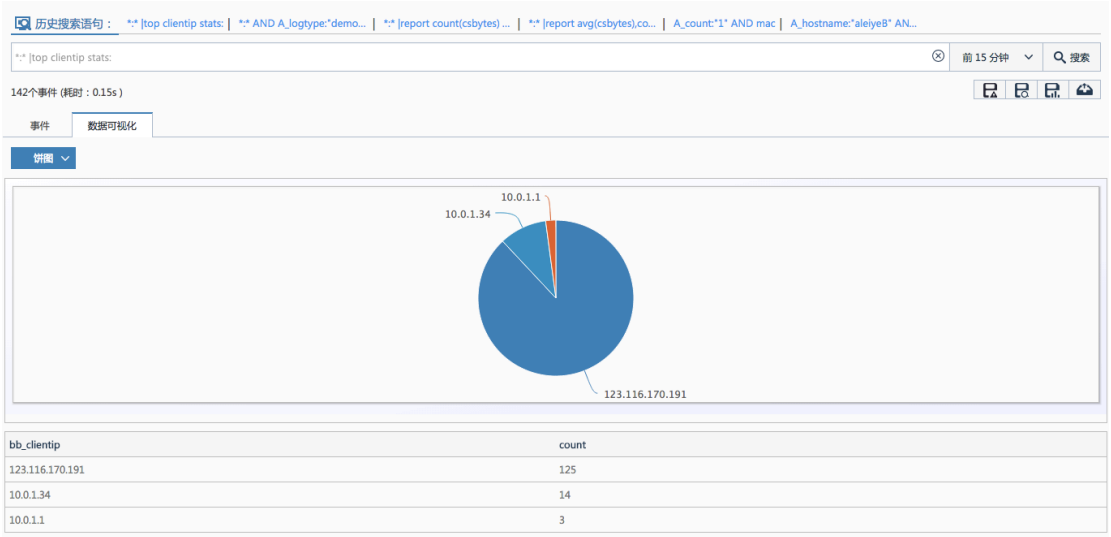


图 4-42

Aleiye 对搜索结果提供告警、保存搜索、保存报表、导出报告或数据等操作。基于搜索语句，可以将结果进行多种形式的保存。

告警：点击告警图标，输入查询语句，设置周期和告警阀值等信息可保存至告警；基于搜索结果，只能保存成计划告警，即周期性执行语句，并到达预设的阈值进行告警操作。

保存报表：将本次搜索保存为报表，可在报表页面查看或编辑已保存的报表；如果搜索语句中，包含报表命令、eval 和 SQL 语句的情况下，是可以将结果保存成报表。

保存搜索：常用的搜索语句和条件可以通过辅助搜索功能进行存储，方便用户快速搜索； 任何搜索命令都可以进行保存。

导出：搜索结果可以导出 CSV 格式的文件；

4.6 SQL 搜索

Aleiye 支持基于 SQL 语句进行数据搜索，当前 ALeiye 支持 SQL2003 标准，除搜索语法不一样外，SQL 搜索也需要确认数据类型，该数据类型可以定义为数据库中表名称，在数据库对接功能说明中提到，系统会将导入的表名称前加上 DB_URL 别名_表明。 原语搜索主要针对近 7 天或 30 天的数据进行实时搜索，而 SQL 搜索增基于历史数据进行大数据量的离线搜索。

点击搜索框打开辅助搜索页面，辅助搜索提供保存搜索条件和保存报表，可快速帮助用户进行搜索操作。如图 4-43。

保存的搜索条件：搜索条件可以进行保存，并在该功能中进行快捷操作。

保存的报表：显示通过搜索结果保存的报表，点击报表可重现报表结果。

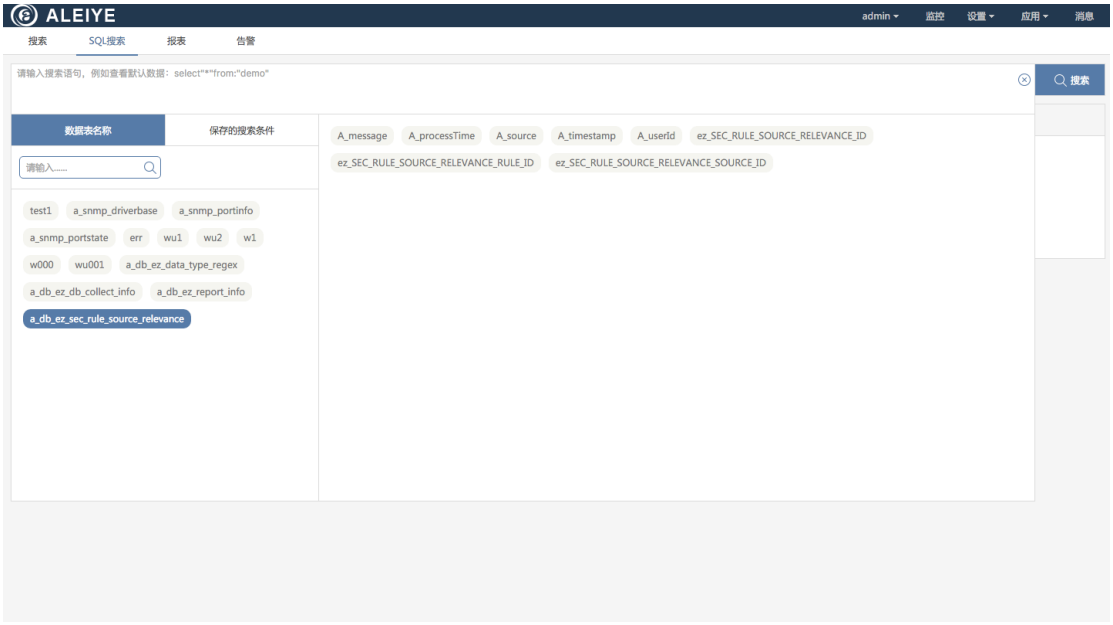


图 4-43

SQL 搜索的结果会议数据透视的形式进行展示，可以根据结果中的字段通过拖拽的形式任意组合，并可以将结果保存成报表。如图 4-44。

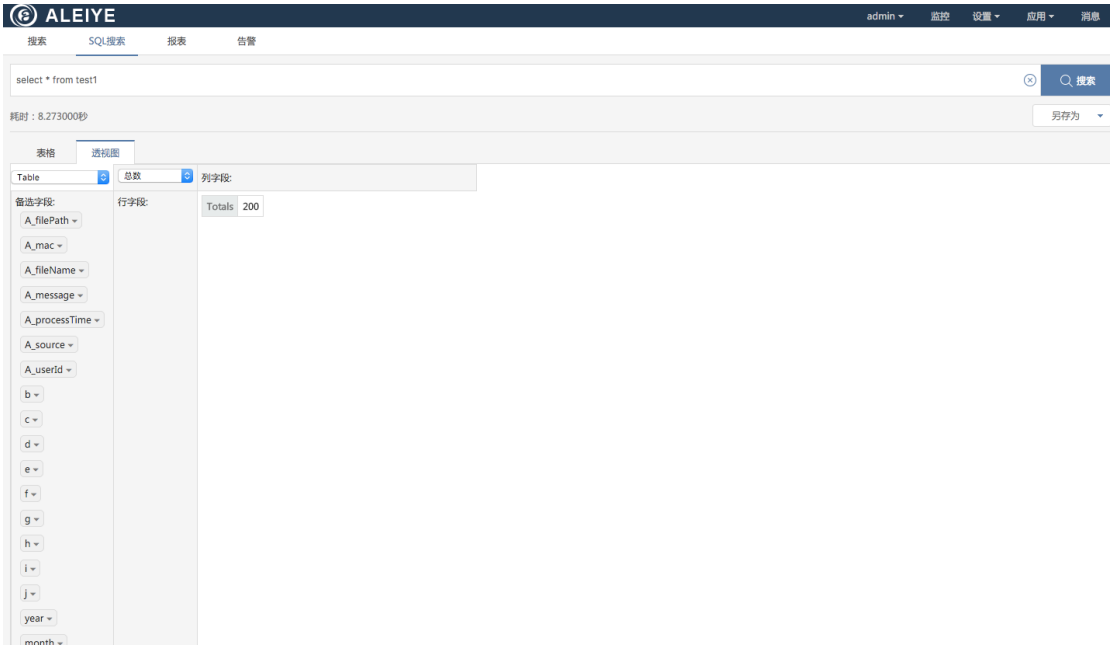
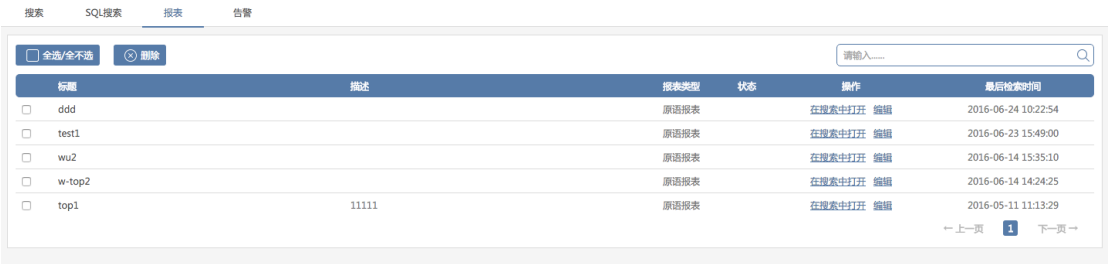


图 4-44

4.7 报表

针对原语搜索和 SQL 搜索的结果，进行报表保存，并可以通过报表页面对所有报表信息进行信息回溯、编辑和删除等功能。

在搜索中打开：在操作功能栏中，点击在搜索中打开，将会依照报表中的搜索语句和数据范围在搜索中进行执行，并展现结果，如图 4-36。



编辑：对报表进行修改操作，可以修改报表名称和描述，搜索语句和数据范围则不可以进行修改。如图 4-37。

4.8 告警

Aleiye 中的告警分为两种类型，一种为计划告警，计划告警主要针对搜索结果，针对搜索结果进行周期性执行，并超出预设阈值后，便进行告警操作；第二种为实时告警，针对数据类型中的字段规则进行自定义规则配置，触发条件后及执行告警操作。

告警管理页面中，可以对告警进行增删改查的操作，并可以查看所有告警记录，如果创建计划告警，必须通过搜索条件进行添加，如果是实时告警，可以在管理页面进行自主添加操作。如图 4-38。

4.8.1 计划告警

计划告警会根据搜索语句进行周期执行，并超过阀值后进行告警操作，如图4-39。创建好的计划告警可以通过邮件的方式或者管理页面中查看功能查看每条告警的具体详细信息。

计划：告警执行周期，分别为每小时执行、每天执行、每周执行、每月执行和 cron 执行，cron 执行为用户自定义执行周期。

搜索语句：通过搜索语句进行设定告警对象，例如搜索语句录入response:"304",则对状态码为 304 的情况进行设定告警条件。

时间：该时间为数据的时间范围，即上述的搜索语句的时间范围。

触发条件：设定启动告警的条件，设定方式分别为等于、大于、大于等于、小于、小于等于和不等于。例如可设定触发条件为“等于 10 或大于等于 9”。

通知方式：Aleiye 支持邮件通知，可以通过邮件的方式通知告警信息。

4.8.2 实时告警

实时告警针对数据类型中的字段进行规则配置，配置规则一旦超出预设的阀值，便进行告警，创建好的告警信息可以进行编辑和删除操作，通过邮件或页面中的查看功能，可以对告警明细进行追溯。如图 4-40。

时间窗口：时间窗口预设的值包括近 1 分钟、5 分钟、10 分钟、15 分钟、30 分钟，该时间即为告警的时间范围。

数据类型：可以选择需要设置告警的数据类型，并展示出该数据类型中的所有字段，方便选择字段间的规则。

过滤条件：“且”“或”代表字段间的规则关系，也可以将规则生成规则表达式，从而实现自定义规则。

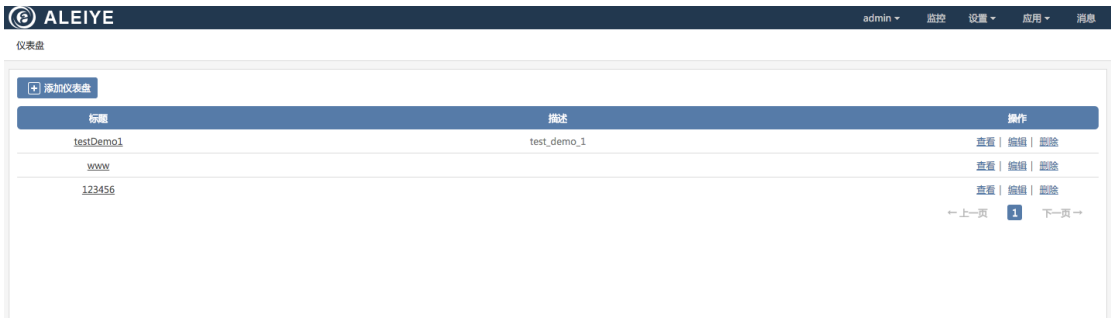
触发条件：配置好上述的规则后，以规则为单位配置该规则的触发条件。

通知方式：与计划告警一致，支持邮件的方式进行告警通知。

4.9 仪表盘

仪表盘可以将多张创建好的报表图形进行统一展示，以便进行实时监测，并可以周期性执行报表。仪表盘功能可以添加到首页，在首页进行关键指标的实时监控。

仪表盘配置：仪表盘中数据来源主要针对报表，当保存好的报表，都可以同步导入到仪表盘中，除此之外，可以根据搜索语句，自定义新的仪表盘。详见图4-41。

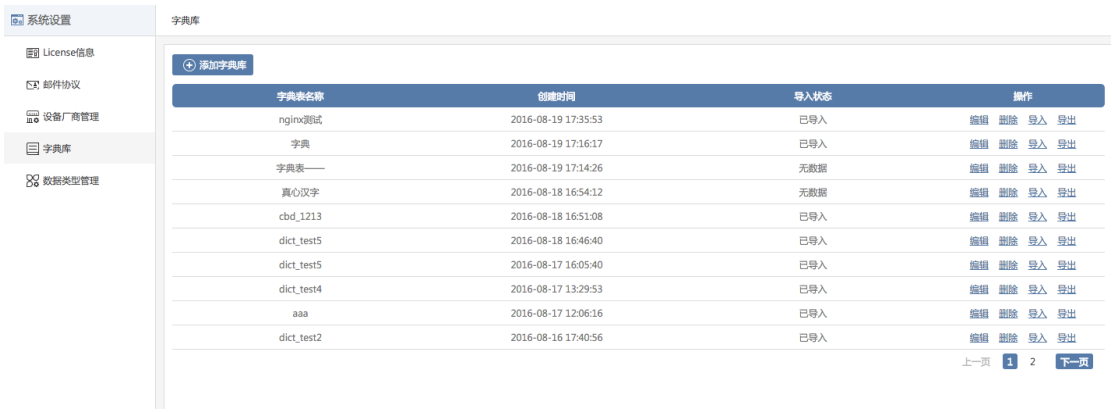


4.10 字典库

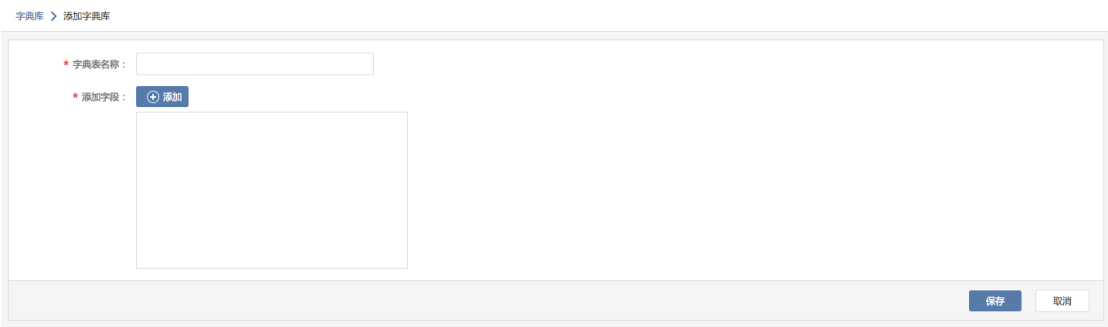
可以在 Aleiye 平台中，创建字典库，通过创建好的字典库，可以在数据入库前进行关联操作，从而实现对数据进行补全的功能。

创建字典库分为两种模式，一种是导入，需要按照平台的字典表结构，导入准备好的字典库，也可以通过页面操作逐条添加字典信息。

字典库的入口在设置-系统设置-字典库，进入字典表管理页面后，可以对创建的字典表进行增删改查的操作。



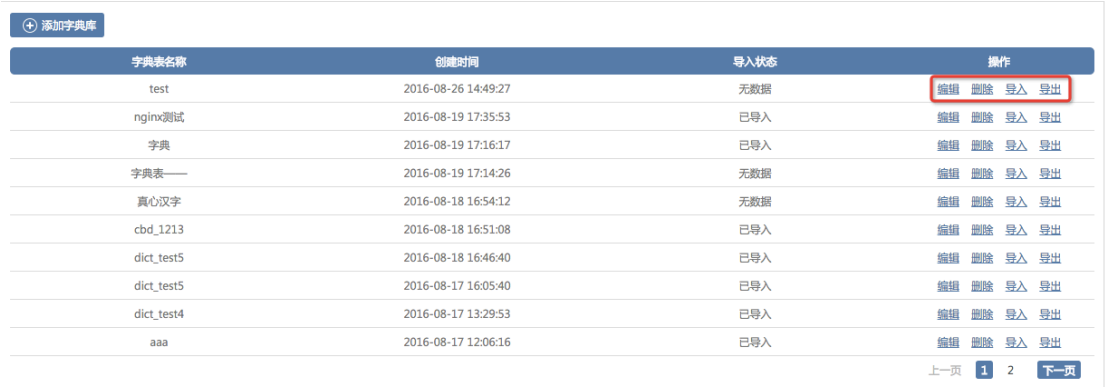
在创建字典表前，需要先创建字典表结构，创建好结构后，可以在该结构的基础上，添加字典信息，字典表结构创建好后是不可以进行修改的。通过字典表管理页面中的添加字典库进入到表结构创建页面。



创建字典表结构的配置页面共有两个配置项，分别是字典表名称和字段。

添加字段：该配置项主要是创建表结构的构成，字典表结构主要是 key、value 的形式，通过配置页面，可以创建多个 key，但必须有一个 key 为主键，可以在配置页面中，自定义主键字段。在创建 key 中，必须选择该字段的字段类型。创建好的字典表结构保存到字典库管理页面，并可以对表结构添加字典信息。

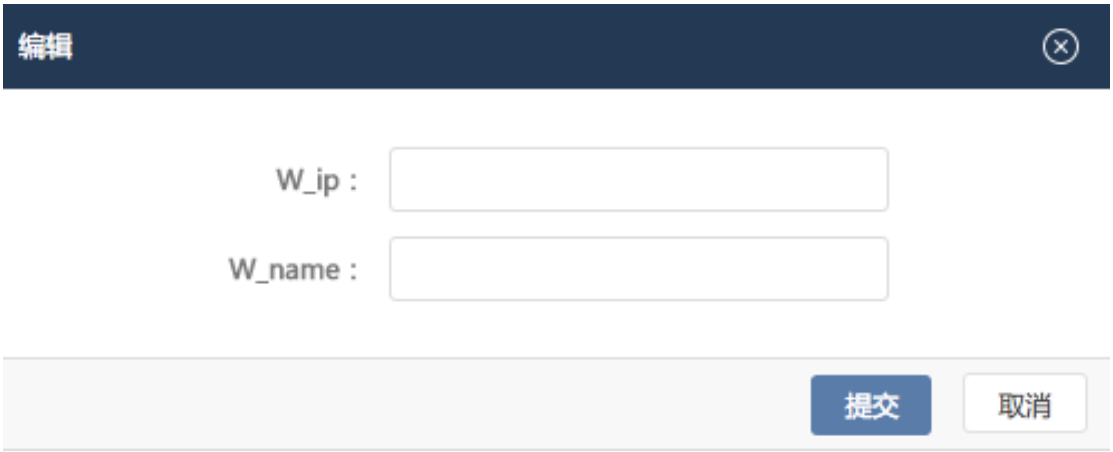
当创建好字典表结构后，便可以在该结构的基础上，添加字典信息了。



在字典库管理页面，可以看到新创建的字典表结构，在操作栏中也可以看到，针对表结构可以执行编辑、删除、导入和导出四个个功能。通过操作中功能，我们可以实现手动添加和批量导入这两种方式添加字典信息。

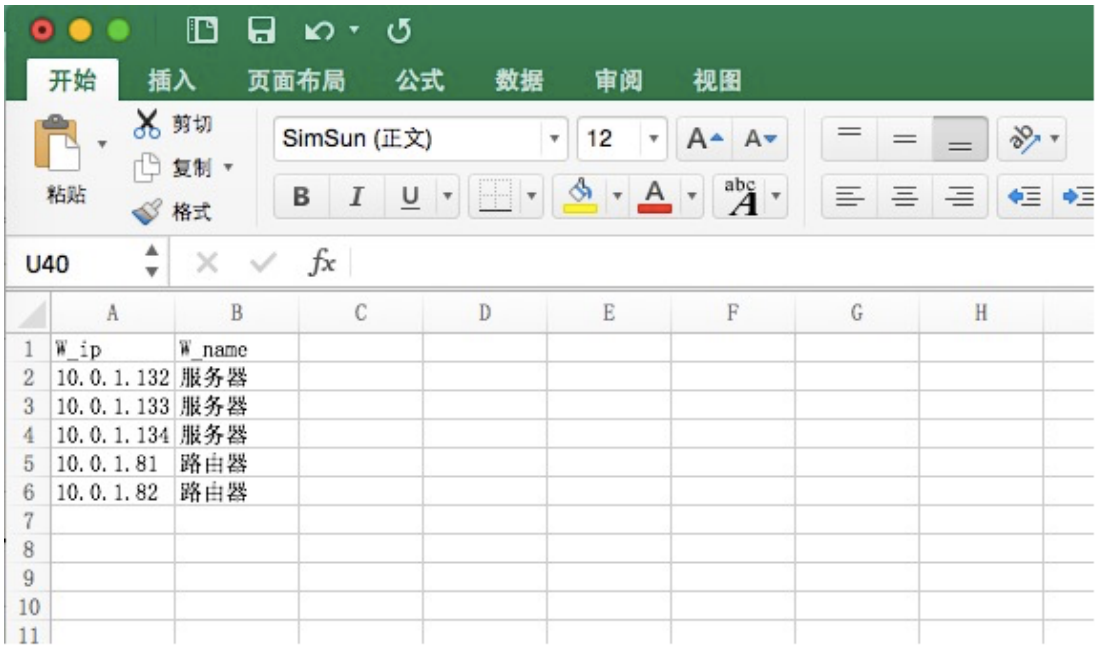
手动添加：手动添加字典信息，是通过页面的交互从而在表结构的基础上添加字典表信息，点击编辑按钮，进入创建好的字典表结构。

进入编辑页面，可以看到，之前添加的字段，已经生成 key value 的形式，通过添加按钮，便可以对应所创建的 key 添加 value 值。同时，我们可以对添加好的 value 值进行二次编辑和删除的操作。

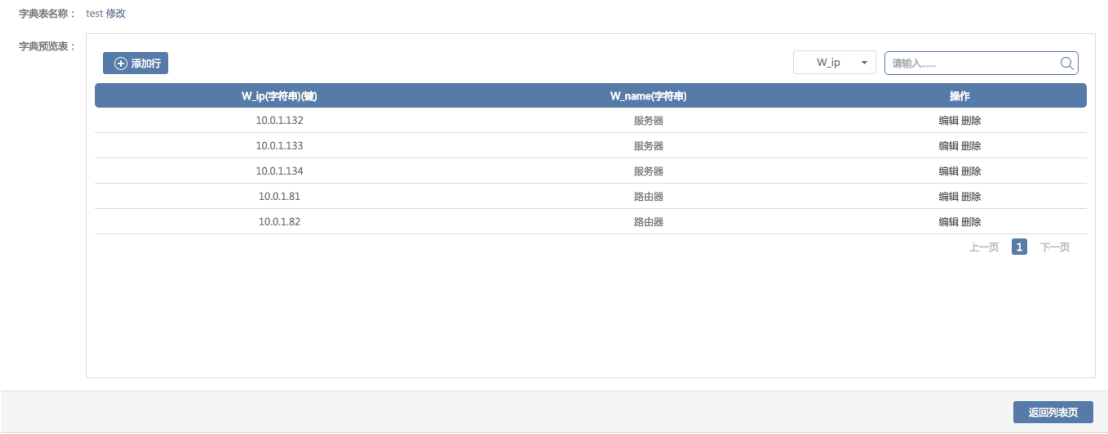


同时，我们可以对添加好的 value 值进行二次编辑和删除的操作。该功能适合在原有的字典库的基础上增加少量的字典信息。如果需要添加大量的字典信息的话，可以通过批量导入的功能。

批量导入：批量导入字典信息，需要现在从页面中下载字典表结构，系统会以 csv 格式将字典表结构导出，只需要在导出的文件中进行字典信息编辑，最后将编辑好的文件导入到平台中即可实现字典表信息批量添加。



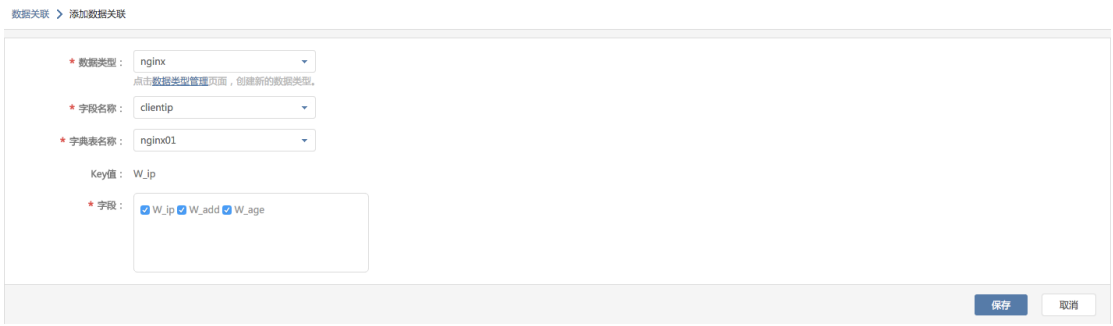
编辑好的字典表通过页面中导入功能，导入到平台中，之后，点击编辑便可以看到批量导入的字典表信息。



4.11 字典表关联

在创建完字典表后，可以通过与字典表进行关联操作，从而实现数据补全的操作。关联操作入口在首页左侧导航中。

点击字典表关联，进入关联管理页面，字典表的关联对象是数据类型，所以在做关联操作之前，需要创建好数据类型，当被关联的数据类型已经被引用的情况下，依然可以执行关联操作，但是历史数据是无法补全关联字段，只有在关联操作完成之后的数据是可以补全相应的关联字段。点击添加关联关系按钮，进入配置页面。



数据类型：数据类型为字典表的关联对象，一旦与某数据类型关联上后，会在该数据类型中补全字典表中的字段。

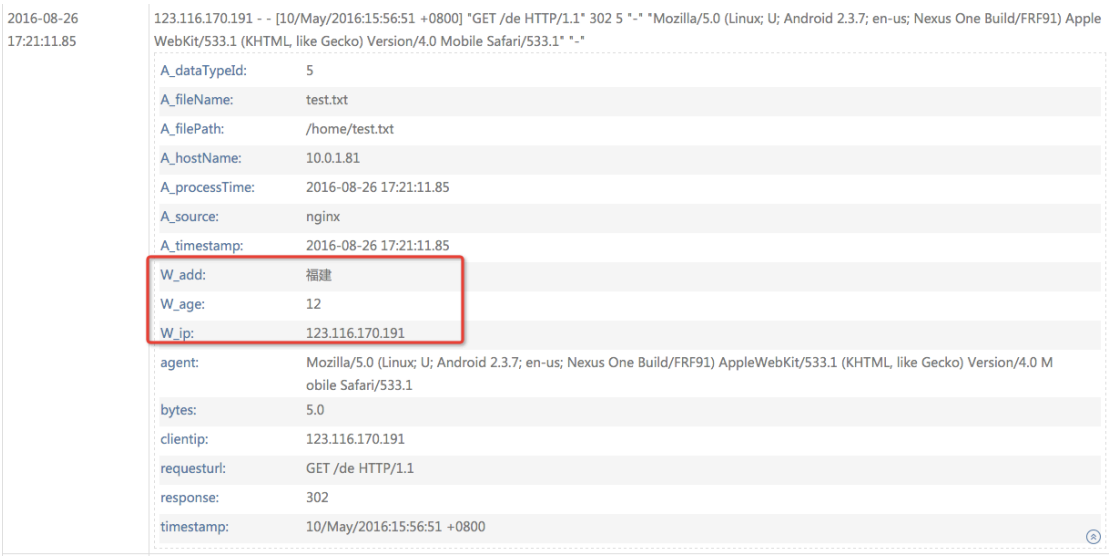
字段名称：选中数据类型后，可以关联出该数据类型下的所有字段，需要选择一个字段作为关联字段，该字段会与字典中 key 进行关联匹配，一旦匹配上后，会在数据类型中不上其所选字段。

字典表名称：上述是对数据类型和数据类型的关联字段选择，该配置项开始选择你要关联的字典表名称。

Key 值：当选中字典表后，会将该字典表中的 key 值进行展示，以便查看关联字段。

字段：选中字典表后，会把该字典表中的所有字段进行展示，你可以选择关联后补全哪些字段。

当所有配置完成后，点击保存，即将该数据类型和字典表的关联关系配置完毕，一旦该数据类型的数据采集上来后，会按照上述的配置内容进行关联，并最终在搜索中可以展示



数据类型与字典表关联上之后，会在搜索中查看到，关联上的字段系统会在该字段增加“W”的前缀，以便做区分，关联补全的字段也可以通过报表等命令机型统计分析。

4.12 数据挖掘

Aleiye 平台中内置四种数据挖掘算法，包括：直线型预测、曲线型预测、异常点监测和异常模式关联。

4.12.1 算法配置说明

时间选择：首先需要选择你要分析的数据的时间点，如果数据为历史数据，则需要选择相应的绝对时间，例如数据源中的时间为 6 月 1 日，则需要在时间范围中选择 6 月 1 日零点到 6 月 1 日 23 点 59 分 59 秒。如果数据是实时的，那可以选择相对时间范围，例如今天、一周迄今等。

数据类型：算法的数据来源是数据类型，在做数据挖掘前，要创建好相应的数据类型，当确认时间范围后，该时间内的所有数据都可以进行筛选。

过滤条件：可以通过检索语句对选中的数据类型进行过滤，如样例数据中，包含四个店铺的访问量，可以通过搜索语句对不需要的店铺数据进行过滤，具体语句如：name:” 宾帅”,通过该语句，数据类型中除了“宾帅”店铺以外的数据全部过滤掉。我们就可以对“宾帅”店铺的数据进行预测。

```
时间,id,name,访问量
2016-06-01 08:00:00,70988183,宾帅,6790
2016-06-01 08:00:00,282546443,轻骑者,2831
2016-06-01 08:00:00,266280753,欧米芝,15438
2016-06-01 08:00:00,31356518,薇池,3023
```

字段选择：选中数据类型后，会关联出相应的字段，曲线预测只能选择一个字段，且该字段类型必须为数字型，该字段即最终预测的指标。

时间区间：时间区间即统计结果的时间粒度，当前支持分钟、小时和天三个粒度。

附件

一、 cron 表达式详解

Cron 表达式是一个字符串，字符串以 5 或 6 个空格隔开，分为 6 或 7 个域，每一个域代表一个含义，Cron 有如下两种语法格式：

Seconds Minutes Hours DayofMonth Month DayofWeek Year 或

Seconds Minutes Hours DayofMonth Month DayofWeek

每一个域可出现的字符如下：

Seconds:可出现", - * /"四个字符，有效范围为 0-59 的整数

Minutes:可出现", - * /"四个字符，有效范围为 0-59 的整数

Hours:可出现", - * /"四个字符，有效范围为 0-23 的整数

DayofMonth:可出现", - * / ? L W C"八个字符，有效范围为 0-31 的整数

Month:可出现", - * /"四个字符，有效范围为 1-12 的整数或 JAN-DEc

DayofWeek:可出现", - * / ? L C #"四个字符，有效范围为 1-7 的整数或 SUN-SAT 两个范围。1 表示星期天，2 表示星期一， 依次类推

Year:可出现", - * /"四个字符，有效范围为 1970-2099 年

每一个域都使用数字，但还可以出现如下特殊字符，它们的含义是：

(1)*：表示匹配该域的任意值，假如在 Minutes 域使用*, 即表示每分钟都会触发事件。

(2)?:只能用在 DayofMonth 和 DayofWeek 两个域。它也匹配域的任意值，但实际不会。因为 DayofMonth 和 DayofWeek 会相互影响。例如想在每月的 20 日触发调度，不管 20 日到底是星期几，则只能使用如下写法： 13 13 15 20 * ?, 其中最后一位只能用？，而不能使用*，如果使用*表示不管星期几都会触发，实际上并不是这样。

(3)-:表示范围，例如在 Minutes 域使用 5-20，表示从 5 分到 20 分钟每分钟触发一次

(4)/：表示起始时间开始触发，然后每隔固定时间触发一次，例如在 Minutes 域使用 5/20,则意味着 5 分钟触发一次，而 25，45 等分别触发一次.

(5),:表示列出枚举值值。例如：在 Minutes 域使用 5,20，则意味着在 5 和 20 分每分钟触发一次。

(6)L:表示最后，只能出现在 DayofWeek 和 DayofMonth 域，如果在 DayofWeek 域使用 5L,意味着在最后的一个星期四触发。

(7)W:表示有效工作日(周一到周五),只能出现在 DayofMonth 域，系统将在离指定日期的最近的有效工作日触发事件。例如：在 DayofMonth 使用 5W，如果 5 日是星期六，则将在最近的工作日：星期五，即 4 日触发。如果 5 日是星期天，则在 6 日(周一)触发；如果 5 日在星期一到星期五中的一天，则就在 5 日触发。另外一点，W 的最近寻找不会跨过月份

(8)LW:这两个字符可以连用，表示在某个月最后一个工作日，即最后一个星期五。

(9)#:用于确定每个月第几个星期几，只能出现在 DayofMonth 域。例如在 4#2，表示某月的第二个星期三。

举几个例子:

0 0 2 1 * ? * 表示在每月的 1 日的凌晨 2 点调度任务

0 15 10 ? * MON-FRI 表示周一到周五每天上午 10：15 执行作业

0 15 10 ? 6L 2002-2006 表示 2002-2006 年的每个月的最后一个星期五上午 10:15 执行作

一个 cron 表达式有至少 6 个（也可能 7 个）有空格分隔的时间元素。

按顺序依次为

秒（0~59）

分钟（0~59）

小时（0~23）

天（月）（0~31，但是你需要考虑你月的天数）

月（0~11）

天（星期）（1~7 1=SUN 或 SUN，MON，TUE，WED，THU，FRI，SAT）

年份（1970－2099）

其中每个元素可以是一个值(如 6),一个连续区间(9-12),一个间隔时间(8-18/4)(/表示每隔 4 小时),一个列表(1,3,5),通配符。由于"月份中的日期"和"星期中的日期"这两个元素互斥的,必须要对其中一个设置?

0 0 10,14,16 * * ? 每天上午 10 点，下午 2 点，4 点

0 0/30 9-17 * * ? 朝九晚五工作时间内每半小时

0 0 12 ? * WED 表示每个星期三中午 12 点

"0 0 12 * * ?" 每天中午 12 点触发

"0 15 10 ? * *" 每天上午 10:15 触发

"0 15 10 * * ?" 每天上午 10:15 触发

"0 15 10 * * ? *" 每天上午 10:15 触发

"0 15 10 * * ? 2005" 2005 年的每天上午 10:15 触发

"0 * 14 * * ?" 在每天下午 2 点到下午 2:59 期间的每 1 分钟触发

"0 0/5 14 * * ?" 在每天下午 2 点到下午 2:55 期间的每 5 分钟触发

"0 0/5 14,18 * * ?" 在每天下午 2 点到 2:55 期间和下午 6 点到 6:55 期间的每 5 分钟触发

"0 0-5 14 * * ?" 在每天下午 2 点到下午 2:05 期间的每 1 分钟触发

"0 10,44 14 ? 3 WED" 每年三月的星期三的下午 2:10 和 2:44 触发

"0 15 10 ? * MON-FRI" 周一至周五的上午 10:15 触发

"0 15 10 15 * ?" 每月 15 日上午 10:15 触发

"0 15 10 L * ?" 每月最后一日的上午 10:15 触发

"0 15 10 ? * 6L" 每月的最后一个星期五上午 10:15 触发

"0 15 10 ? * 6L 2002-2005" 2002 年至 2005 年的每月的最后一个星期五上午 10:15 触发

"0 15 10 ? * 6#3" 每月的第三个星期五上午 10:15 触发

有些子表达式能包含一些范围或列表

例如：子表达式（天（星期））可以为 “MON-FRI”，“MON，WED，FRI”，“MON-WED,SAT”

“*”字符代表所有可能的值

因此，“*”在子表达式（月）里表示每个月的含义，“*”在子表达式（天（星期））表示星期的每一天

“/”字符用来指定数值的增量

例如：在子表达式（分钟）里的“0/15”表示从第 0 分钟开始，每 15 分钟

在子表达式（分钟）里的“3/20”表示从第 3 分钟开始，每 20 分钟（它和“3，23，43”）的含义一样

“？”字符仅被用于天（月）和天（星期）两个子表达式，表示不指定值

当 2 个子表达式其中之一被指定了值以后，为了避免冲突，需要将另一个子表达式的值设为“？”

“L” 字符仅被用于天（月）和天（星期）两个子表达式，它是单词“last”的缩写

但是它在两个子表达式里的含义是不同的。

在天（月）子表达式中，“L”表示一个月的最后一天

在天（星期）自表达式中，“L”表示一个星期的最后一天，也就是 SAT

如果在“L”前有具体的内容，它就具有其他的含义了

例如：“6L”表示这个月的倒数第 6 天，“FRIL”表示这个月的最一个星期五

注意：在使用“L”参数时，不要指定列表或范围，因为这会导致问题

字段 允许值 允许的特殊字符

秒 0-59 , - * /

分 0-59 , - * /

小时 0-23 , - * /

日期 1-31 , - * ? / L W C

月份 1-12 或者 JAN-DEC , - * /

星期 1-7 或者 SUN-SAT , - * ? / L C #

年（可选） 留空, 1970-2099 , - * /

ALEIYE

让 | 大 | 数 | 据 | 更 | 简 | 单

公司地址：北京市西城区新街口外大街28号普天德胜大厦A座405

邮　　编：100088

联系电话：010-82053991