数据分析平台-SAAS

v 1.x.x

2020年5月6日

# 版本信息

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **修订时间** | **修订人** | **修订内容** |
| **1.0** | 2020年5月6日 | 边学明 | 应用管理/主界面 |
| **1.1** | 2020年5月7日 | 边学明 | 概览 |
| **1.2** | 2020年5月13日 | 边学明 | 元数据 |
| **1.3** | 2020年5月26日 | 边学明 | 细化入库校验、事件结构 |
|  |  |  |  |

目录

# 内容

用户群组

# 内容

应用管理

应用中的数据与配置相互独立。每个用户可能创建多个应用，多用户可以同享同一个应用数据及配置。



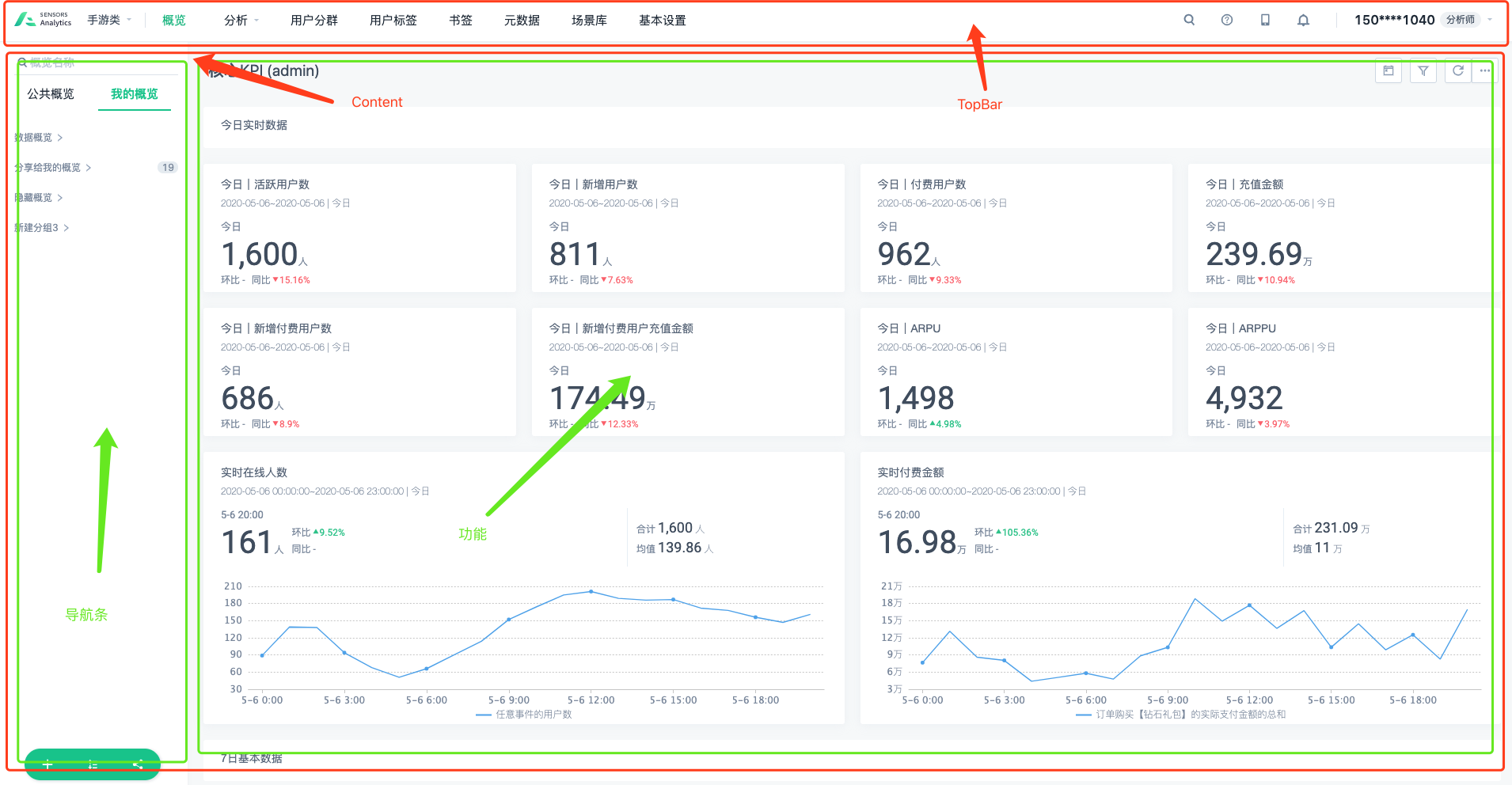
【创建应用】

# 内容

项目主页

用于管理当前项目（应用）的入库数据和配置。所有分析的依据都是以当前项目为基础进行操作的。

界面分为【导航栏】【内容栏】两大块组成，【内容栏】又细分为【左侧功能栏】和【右侧显示栏】。【内容栏】的显示内容会根据【导航栏】进行变化，有些会有更具体的细分功能，有些会全屏显示。



【导航栏-功能】大功能分类，默认选中第一项。如果有子分类会以浮动菜单的形式展现。

【左侧导航栏】显示选中大功能的子分类或数据列表（有的功能会没有左侧导航栏），默认选中第一项。【右侧内容栏】根据选中的【左侧导航栏】变化。

（状态展示）

# 内容

概览

保存常用的分析结果，通过概览方便查看的数据。同时用户可以将自己的概览共享给团队其他成员，来分享分析结果。需要注意的是，数据结果将按照对方自己的数据权限范围来显示结果。

1. 概览分类
   1. **我的概览**

我的概览主要用于使用成员，自己将常用的分析结果保存起来，形成便于查看的数据概览。同时用户可以将自己的概览共享给团队其他成员，来分享分析结果。需要注意的是，数据结果将按照对方自己的数据权限范围来显示结果。

* 1. **公共概要**

公共概览主要用于管理团队需要共同关注的重要概览，打造团队的数据公共空间，由具有「管理公共概览」权限的成员共同维护。

在公共概览可以进行创建、复制、排序等基础操作，为了使用的顺畅性，所有的功能操作方式和「我的概览」一致。

1. 概念介绍
   1. **概览**

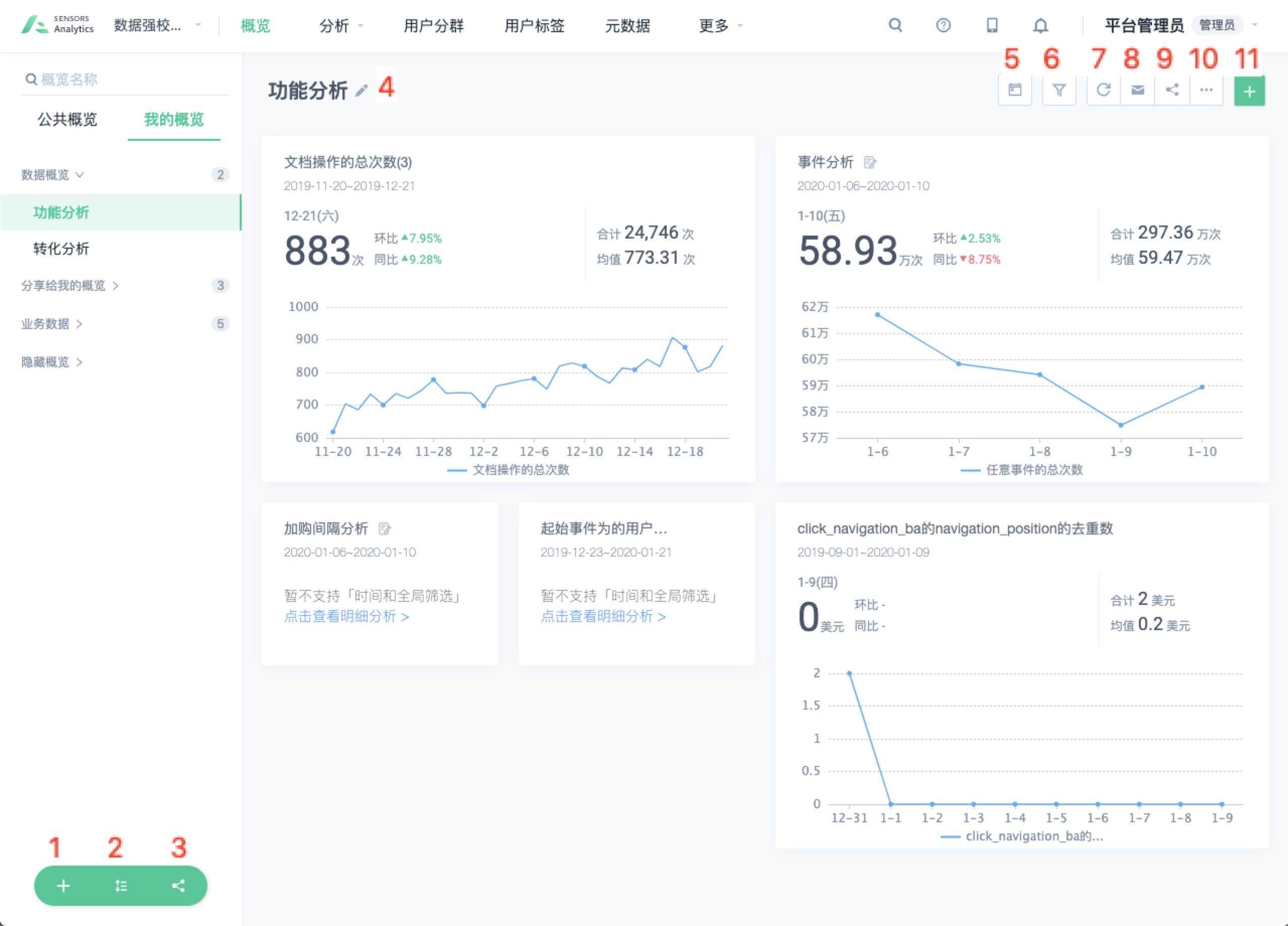
用于管理和查看多个【分析】的容器，除了对分组进行“创建”、“删除”等操作外，还可以为分组“增加”、“删除”和“管理”分组中包含的分析组件，将多个分析组件打包成一组方便相关业务的查看。

* 1. **分组**

用于管理【分组】的容器，可以对概览进行“创建”、“管理”、“删除”等相关操作。通过概览可以更直观、更方便的对分组进行查看。

1. 概览/分组操作

以【我的概览】为例进行基本操作使用的介绍。



1. 新建概览分组与概览
2. 概览排序管理：支持修改分组名称
3. 概览批量共享设置、删除的管理入口
4. 修改当前概览名称
5. 全局筛选时间窗
6. 全局筛选事件属性和用户属性
7. 刷新当前的计算数据（仅刷新失效数据，或刷新全量数据）
8. 概览邮件发送设置
9. 共享设置
10. 更多操作：~~性能设置、全屏模式、停止更新、复制概览、~~删除概览
11. 在概览中添加数据信息：从书签添加已有内容、新建组件、新建描述
    1. **新建【概览】【分组】**

在【公共概览】和【我的概览】的底部操作，点击「+」可以在当前分类下创建新的概览或者分组。

只有具有「公共概览管理」权限的成员才可以编辑此页签，为了方便操作，~~可以在各个分析模型中分析后进行保存时，可选择此分析结果保存为书签后，添加到「公共概览」或「我的概览」中。~~

* 1. **排序**

在【公共概览】和【我的概览】的底部操作，点击中间的「排序按钮」后可进行当前概览的排序。

* 为了你能便捷的排版概览，可多选概览后进行批量移动的操作。
* 在此弹窗中，支持鼠标按住「概览」或「分组」可直接进行排序拖动。
* 此外，鼠标聚焦住具体「概览」或「分组」还可进行概览、分组时，自已经创建的会出现，“重命名”、“删除”和“移动到”快捷操作。
  1. **概览批量共享设置、删除管理**

选择【公共概览】或【我的概览】，在其左侧导航的「最右侧的按钮」点击后可分别进入对应的【概览管理】页面。【概览管理】将会占满【内容栏】，可以进行单选或是多选对【概览】或【分组】进行下面的操作。

* 支持搜索概览名称关键字进行检索
* 支持进行「分享设置」，通过此配置设置当前概览的可见范围：「仅有管理公共概览权限的成员可见」、「全体成员」、「指定成员」；在「指定成员」中可以设置某些成员，或某些角色动态配置，或某些职务的动态配置。注：如果当前项目没有「职务信息」此处将不显示。你可以前往「成员与角色」-「成员管理」-「职务管理」中添加符合需要的。随后是此职务的成员，均可按照此配置查看到对应的概览。
* 编辑操作，支持在此修改概览的名称、与分组设置
* 删除操作，删除后，以前可查看此概览的成员，将不可查看；
* 停止分享，操作后「仅有管理公共概览权限」的成员方可查看。

1. 内容的显示
2. **修改当前分组名称**

点击编辑按钮后，可直接修改概览名称，随后按 Enter 键或点击非输入区域将完成重命名的保存。

1. **时间范围**

修改时间后，将对此概览内的所有分析结果生效。当切换概览或页面后，此时间条件不支持保留

1. **晒选条件**
2. **刷新**

根据选择的条件再次刷新分析组件。

1. **邮件发送设置**
2. **共享设置**

可将自己的概览共享给其他成员查看，共享成功后将在通知助手中通知对方，并自动添加到对方的「我的概览」-「分享给我的概览」中。当你修改了此概览内容，对方查看时将同步更新。

可将自己的概览共享给全部成员，或指定的角色、职务、或具体成员。

1. **更多操作：删除**

删除当前分组。

# 内容

分析

# 内容

元数据管理<ETL清洗、验证>

可以统一管理所追踪数据元信息和配置。用户和事件属性的名称是全局唯一，且类型一致的。

通过「数据强校验模式」可以有效的提升数据的准确性，从根源解决错误数据上报入库的问题。开启此模式后，将需要采集的事件与属性更新到项目中后，后续所有上报入库的数据均需要比对已配置的事件与属性类型进行校验。符合要求的数据方可入库，不符合要求的数据将无法入库。

使用「自由数据上报模式」则可以让企业快速完成数据采集的工作，但是数据质量需要完全依赖研发者的埋点代码的准确性。在此模式下，可以随时上报任何事件与属性。后续持续上报的属性相关数据，将以首次入库的属性类型和属性名进行对比校验。如未匹配到库里已有的信息，那么将视为一个全新的「事件」、「事件属性」、「用户属性」入库到当前项目中。

**注意：属性的名称是全局唯一的，且类型是一致的。如果有两个事件，比如注册和登录中，都有属性 username 的话，必须保证这个 username 的类型是一致的。**

1. 数据格式
2. **日志**

日志文件是一行一个 JSON，物理上对应一条数据，逻辑上对应一个描述了用户行为的事件，或是描述一个或多个用户属性的 Event 操作。

{

"distinct\_id": "123456",

"time": 1434556935000,

~~"type": "track",~~

"event": "ViewProduct",

~~"project": "ebiz\_test",~~

~~"time\_free": true,~~

~~"properties": {~~

~~"$is\_login\_id":true,~~

~~"$app\_version":"1.3",~~

~~"$wifi":true,~~

~~"$ip":"180.79.35.65",~~

~~"$province":"湖南",~~

~~"$city":"长沙",~~

~~"$user\_agent":"Mozilla/5.0",~~

~~"$screen\_width":320,~~

~~"product\_classify":"水果",~~

~~"product\_price":14.0~~

~~}~~

}

对于上述字段的说明如下：

* distinct\_id：**类型是字符串**，对用户的标识，对未登录用户，可以填充设备标识、CookieID 等，对于登录用户，则应该填充注册账号；如果是一个群体事件可以省略此字段；
* time：**类型是数值**，事件发生的实际时间戳，精确到毫秒；
* ~~type：track 表明是记录一个 Event；~~
* event：事件名，需是合法的变量名，即不能以数字开头，且只包含：大小写字母、数字、下划线和 $，其中以 $ 开头的表明是系统的保留字段，自定义事件名请不要以 $ 开头，且 event 字段长度最大为 100；
* project：这条数据所属项目名，若不指定该参数，则需要使用该字段时取值 **default**，即默认项目。指定的项目必须是系统中已经存在的项目，否则这条数据将无效，更多项目相关请参见[多项目](https://manual.sensorsdata.cn/sa/latest/tech_knowledge_more-1573783.html)；
* time\_free：可选字段，表示不根据事件发生时间过滤该事件。**只要出现 time\_free 这个 key 且 value 不为 null，将不再校验 time 是否在允许导入的时间范围内**。导入历史数据时可能会用到该字段；
* ~~properties：这个 Event 的具体属性，以 dict 的形式存在。其中以$开头的表明是系统的保留字段，它的类型和中文名已经预先定义好了。自定义属性名需要是合法的变量名，不能以数字开头，且只包含：大小写字母、数字、下划线，自定义属性不能以 $ 开头；同一个名称的 property，在不同 event 中，必须保持一致的定义和类型；同一个名称的 property 大小写不可以相同，如果已经存在小写属性就不可再导入对应大写属性（比如元数据中有 abc 属性名，不能再传 ABC,Abc 等属性名），否则数据会校验失败不入库。~~

1. **Profile**

Profile 为用户存档数据，可以通过远端传入，也可以通过导入存档。

{

"distinct\_id": "12345",

"type": "profile\_set",

"time": 1435290195610,

"project": "ebiz\_test",

~~"properties": {~~

~~"$province":"湖南",~~

~~"FavoriteFruits": ["苹果","香蕉","芒果"],~~

~~"Age":33,~~

~~"$city":"长沙",~~

~~"IncomeLevel": "3000~5000",~~

~~"$name": "小明",~~

~~"Gender":"男",~~

~~"$signup\_time": "2015-06-26 11:43:15.610"~~

~~}~~

}

1. **Type 属性类型**

系统以 JSON 数据类型为基础再加以额外限制，定义了若干种数据类型，但 不与 JSON 类型完全等价，详见后文属性类型说明。

|  |  |
| --- | --- |
| 消息类型 | 目标数据表 |
| track | events |
| profile\_\* | users |

1. **数据属性类型**

一个属性的类型由首次导入时的类型决定，元数据-事件属性/用户属性页面展示的属性类型即为首次导入时的类型，后续导入数据时若类型和元数据中展示的类型不符，则尝试对数据进行类型转换，若无法转换或转换失败则 输入数据会被整条拒绝 。尝试进行的类型转换如下（空格不进行转换）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据类型 | 中文名 | 基础 JSON 类型 | 额外限制 | 示例 | 说明 |
| NUMBER | 数值型 | number | -9E15 到 9E15 小数点后最多保留3位 | 12 或 12.0 |  |
| BOOL | 布尔值 | bool | 无 | true 或 false |  |
| STRING | 字符串 | string | 使用 UTF-8 编码后最大长度 1024 字节，如需调整最大长度可以联系我们 | "SensorsData" |  |
| LIST | 字符串数组 | list | 默认是字符串元素的数组（传入的字符串不会去重），最大元素个数为 500，其中每个元素使用 UTF-8 编码后最大长度 255 字节；注意：1.12 之前的版本，List 是集合，即在入库时会进行重排序和去重。如果需要调整 List 具体是数组还是集合，请联系神策技术支持。若 append 导致超过最大元素个数时，新入库的元素会淘汰最早入库的元素。 | ["橘子","西瓜"] |  |
| DATETIME | 日期时间 | string | yyyy-MM-dd HH:mm:ss.SSS 或 yyyy-MM-dd HH:mm:ss 或 yyyy-mm-dd （时分秒按00:00:00处理）， 建议使用第一种，其中 SSS 为毫秒；年取值范围是 [1900, 2099] | "2015-06-19 17:51:21.234" "2015-06-19 17:51:21" "2015-06-19" | 有些语言的时间格式化库不直接提供毫秒格式化，如 Python 提供 us 而不提供 ms，如需自行编写程序生成数据请注意这点。 |

1. **保留字段**

为了保证查询时属性名不与系统变量名冲突，设置如下保留字段，请避免其作为事件名和属性名（properties 中的 key）使用：

date

datetime

distinct\_id

event

events

first\_id

id

original\_id

device\_id

properties

second\_id

time

user\_id

users

1. 入库校验规则设置
   1. **共公设置**

上报事件属性的类型与「事件属性」已有事件的类型不一致时，系统将按规则尝试自动转换，如无法完成则视为异常数据不能入库。

数据入库强校验模式：选中后会强制对数据进行验证。

~~主题设置：设置可读取Kafka的topic, 可通过“，”来设置多个。~~

解析格式：kafka 中的数据以什么形式解析，JSON 存的是JSON格式(默认)。任意格以，通过正则解析出来一个key,value结构的Map。分格符，以自定义分格符的一个数组。（可以配置多个，尝试找到一个合适的）

**接收数据的匹配及验证规则:**

**Key 规则：**

属性匹配：选择一个共公字段

正则匹配：通过正则进行匹配

Text: 与输入的字符完全相同

**入库规则：**默认的情况下为没有

可以通过编写正则进行相应的入库的强验证。

**接收用户（设备ID）：device\_id 的赋值规则及验证规则，并通过【设备ID】来确定是否为同一用户。**

**接收用户（用户ID）：distinct\_id 的赋值规则及验证规则。**

* 1. **事件与事件属性的入库设置**

上报用户属性的类型与「用户属性」已有事件的类型不一致时，系统将按规则自动转换，如无法自动转换则视为异常数据不能入库。

接收事件（发生时间）：time 的赋值规则及验证规则。

接收事件（名称）：event 的赋值规则及验证规则。

接收事件（项目ID）：project 的赋值规则及验证规则。

* 1. **用户属性入库设置**

上报用户属性的类型与「用户属性」已有事件的类型不一致时，系统将按规则自动转换，如无法自动转换则视为异常数据不能入库。

接收用户属性（发生时间）：time 的赋值规则及验证规则。

接收用户属性（项目ID）：project 的赋值规则及验证规则。

接收用户属性（type）：暂时先用默认值”profile\_user”。

1. 元事件

分析及展示模型的数据源。元事件可以进行增、删、改操作，并通过管理元事件从而影响数据的入库及分析等操作。

* 1. **概要**
* 可直接通过搜索「事件名」和「显示名」关键字，或筛选「显示状态」、「标签」、「应埋点平台」快速找到你需要的事件。
* 支持对「显示」、「隐藏」、「添加标签」、「修改标签」的批量操作。
* 对于「已入库」的不再需要接收数据的事件，提供了对整个事件的「停止入库」功能。点击停止入库后，后续产生的数据均不再入库；如业务需求发生变化，又需要继续获取此事件的现今上报数据，点击「允许接收」即可，操作后的上报数据将正常入库，历史未入库的数据不会被恢复。
* 管理元事件的时候，点击 icon 后可直接在事件分析中查看此事件近期数据上报的情况。
* 提供一键下载当前项目中已有的「数据埋点采集方案文档」，文档中将包括此项目中所有的事件、事件属性与用户属性。
* 可通过上传 Excel 文件，对项目中的事件进行显示名的重命名。
* 在「数据强校验模式」下，有数据上报的事件才可在分析模型和权限中使用，且不可被删除。对于无数据上报的事件，可直接进行删除操作。
  1. **创建(仅在【数据强校验模式】下可使用)**

填写事件基本信息后，将此事件需要上报的属性选择添加到此事件中，或直接新建所需的属性，创建后会自动生成到事件属性中。对于公共事件属性，可选择性的使用部分内容或使用全部内容，系统默认显示当前项目中所有的公共事件属性。

**事件属性的字段：**

|  |  |
| --- | --- |
| **信息** | **说明** |
| **事件显示名** | 事件在使用过程中的显示名称 |
| **事件名称** | 仅可命名为英文，是事件在系统内的唯一标示，不能以数值、符号开头。 注：校验时，事件名称大小写敏感。 |
| **应埋点平台** | 表明此事件需要埋点上报的端，可多选。 iOS、Android、JS、后台、其他 |
| **标签** | 非必填项，可自定义。 设置后可通过标签进行事件使用的动态授权，便于管理和筛选。 |
| **触发时机** | 描述什么时候触发这个事件的数据上报。便于研发理解，确保上报数据的时机是正确的。例如：是在点击按钮的时候上报，还是已进入页面后即上报。 |
| **备注** | 帮助业务同学更好理解的信息 |
| **事件属性** | 在这个事件中需要采集的具体信息。 可选择项目中已有的，也可直接添加新的属性。 对于从项目中选择的属性，仅可修改其「显示名」 对于新建的暂无数据上报的属性，可修改其类型、属性名和显示名。 |
| **公共事件属性** | 公共事件属性是通过配置后进行统一上报的，无需在每个事件的埋点中进行。 在此可选择当前事件需要的公共事件属性。 |
| **预置事件属性** | 预置事件属性，是提前在 SDK 中预置好的。 如果需要请在埋点设置的时候进行配置。 可手动预置事件属性中选择你需要的。 |

* 1. **批量创建（EXCEL格式在数据模型表中）**

通过的「数据埋点采集方案文档」将要采集的事件、事件属性、用户属性按照模板要求整理好，然后进行上传。可点击「下载数据埋点采集方案文件」进行下载。

* 1. **编辑**

编辑已经创建的事件，“事件名”和“应埋平台”不可修改。各事件属性可以做增删，不可修改事件属性中的内容。

* 1. **删除**

暂时不提供删除功能，可以通过“显示状态”和是否接收来控制。

* 1. **事件详情**

点击具体的事件可查看到，此事件的基本信息，以及所有关联属性的数据上报状态和显示状态。点击对应的属性，可直接查看其详情信息。在此界面对事件进行编辑。

1. 事件属性

元事件的属性字段，每个字段可用于多个不同的元事件中，但每个字段都是唯一的。

1. **概要**

* 提供创建事件属性的功能，并支持通过上传文件批量修改事件属性的显示名。
* 支持直接搜索「属性名」和「显示名」关键字，或筛选「显示状态」、「数据类型」、「上报数据的状态」和「仅查看公共属性」快速找到你需要的事件。
* 支持对选中属性的「显示」和「隐藏」进行批量操作。
* 在「数据强校验模式」下，有数据上报的事件属性才可在分析模型和权限中使用，且不可被删除。对于无数据上报的事件属性，可直接删除。

1. **新建**

除了在元事件中创建新的事件属性外，当发现某些属性是需要单独添加到多个已有事件中进行埋点上报的时候，也可通过手动在线上创建一个事件属性来完成。点击「创建」打开新建页面，选择需要关联的事件后，完成此事件属性的基本信息补充后方可创建成功。

**事件属性的字段：**

|  |  |
| --- | --- |
| 信息 | 说明 |
| **属性显示名** | 属性在使用过程中的显示名称，100 字符以内。 |
| **属性名** | 仅可命名为英文，是事件在系统内的唯一标示。 不能以数值、$ 符号开头，100 字符以内。 注：校验时，属性名大小写不敏感。 |
| **设置为公共属性** | 公共事件属性是通过配置后进行统一上报的，无需在每个事件的埋点中进行。 |
| **将应用到…事件** | 如果当前属性不是公共属性，那么需要选择此属性关联的事件有哪些。 如果是公共属性，那么不需要进行关联设置，系统会默认为后续新建的所有事件都关联此属性。 |
| **数据类型** | 必填单选 NUMBER、BOOL、STRING、DATETIME、LIST |
| **应埋点平台** | 设置此属性为公共属性后，需填写此信息。用来表明此属性需要配置的端，可多选。 iOS、Android、JS、后台、其他 |
| **设置时机** | 设置此属性为公共属性后，需填写此信息。用来描述什么时候需要设置，便于研发理解。 |
| **单位 / 格式** | 统计值的单位，设置后会在分析详情和概览中显示。 |
| **字典** | 当此属性有数据上报后，可在事件属性列表中上传、管理对应的维度字典。 |
| **属性值示例或说明** | 非必填，便于验证埋点时的研发自测对照。 |

1. **编辑**

编辑已经创建的属性，只能修改“属性显示名”和、“字典”、“设置时机”和“说明”，其它不可修改。

1. **删除**

暂时不提供删除功能

1. **属性详情**

点击具体的事件属性，也可查看到，此属性目前关联的事件明细列表及其显示状态。在此界面对事件属性进行编辑。

1. 用户属性（profile）

用户存档的属性字段，可以是第三方传入，也可以通过SDK进行采集。

1. **概要**

* 提供创建用户属性的功能，并支持通过上传文件批量修改事件属性的显示名。
* 支持直接搜索「属性名」和「显示名」关键字，或筛选「显示状态」、「数据类型」和「上报数据的状态」快速找到你需要的用户属性。
* 支持对选中属性的「显示」和「隐藏」进行批量操作。
* 在「数据强校验模式」下，有数据上报的用户属性才可在分析模型和权限中使用，且不可被删除。对于无数据上报的用户属性，可直接删除。

1. **新建**

除了在元事件中通过上传 Excel 文件创建新的用户属性外，如需要单独添一个新的用户属性时，可通过手动在线完成创建。点击「创建」打开新建页面，填写相关信息后方可创建成功。随后埋点采集的用户属性将比对此设置信息进行校验。

**用户属性的字段：**

|  |  |
| --- | --- |
| **信息** | 说明 |
| **属性显示名** | 属性在使用过程中的显示名称，100 字符以内。 |
| **属性名** | 仅可命名为英文，是事件在系统内的唯一标示。 不能以数值、$ 符号开头，100 字符以内。 注：校验时，属性名大小写不敏感。 |
| **数据类型** | 必填单选 NUMBER、BOOL、STRING、DATETIME、LIST |
| **应埋点平台** | 设置此属性为公共属性后，需填写此信息。用来表明此属性需要配置的端，可多选。 iOS、Android、JS、后台、其他 |
| **设置时机** | 设置此属性为公共属性后，需填写此信息。用来描述什么时候需要设置，便于研发理解。 |
| **单位 / 格式** | 统计值的单位，设置后会在分析详情和概览中显示。 |
| **字典** | 当此属性有数据上报后，可在事件属性列表中上传、管理对应的维度字典。 |
| **属性值示例或说明** | 非必填，便于验证埋点时的研发自测对照。 |

1. **属性详情**

点击具体的用户属性，也可查看到，此属性埋点上报的具体信息明细。

# 内容

事件查询组件

通过对应条件查询和展示所有入库的元数据。可以通过查询条件过滤出对应的当前项目的元事件。

1. 分组

为方便查找，可以自定义不同的分组。组内可以任意添加不同的分类，组内不能存在相同的分类，同一种分类可以存在于不同组中。

系统默认存在一个【共享分组】，其他人共享的【分类】显示在这里，并且只能查看，不能编辑及删除等操作。

1. 分类

分类是对应一个【元事件】，分类可以共享给他人，共享后可以显示在他人的系统分组【共享分组】中，共享的分类可以添加到自定义的分组里，自定义分组/分类需要保存。

1. 查询条件

查询条件分为【事件时间】、【主键】和【自定义】。

|  |  |
| --- | --- |
| **事件时间** | 每条【元事件】都会存在【创建时间】的字段。 |
| **主键** | 可以通过多个主键过滤数据据。 |
| **自定义** | 非主键字段可以增加自定义。 |

自定义条件可以没有，由用户自定义增删个数。

自定义-判断条件

|  |  |
| --- | --- |
| **等于** | 字段值与条件完全相同。 |
| **不等于** | 字段值与条件完全不同。 |
| **大于** | 适用于NUMBER和DATETIME型字段。(>) |
| **小于** | 适用于NUMBER和DATETIME型字段。(<) |
| **区间** | 适用于NUMBER型字段。(IN) |
| **有值** | 适用于STRING和LIST型字段。(SUBSTRING/LIKE) |
| **没值** | 适用于STRING和LIST型字段。 |

时间条件-列出所有【元事件】的创建时间在此范围内的记录。

1. 数据展示

根据条件晒选后的数据以表格形式显示，字段以显示名显示，如果没有用字段名显示。

# 内容

# 内容

用户标签管理

<https://manual.sensorsdata.cn/sps/latest/page-7543687.html>

# 内容

书签

# 内容

报告

# 内容

设置

项目限额

# 内容

分析

漏斗分析

漏斗模型主要用于分析一个多步骤中每一步的转化与流失情况

**基本概念**

1. 步骤：由一个 元事件/虚拟事件 加一个或者多个筛选条件组成，表示一个转化流程中的一个关键性的步骤；
2. 时间范围：在界面上选择的时间范围，是指漏斗的第一个步骤发生的时间范围；
3. 窗口期：用户完成漏斗的时间限制，也即只有在这个时间范围内，用户从第一个步骤，行进到最后一个步骤，才能被视为一次成功的转化；

**计算规则**

假设一个漏斗中包含了 A、B、C、D、E 五个步骤，选择的时间范围是 2015 年 1 月 1 日到 2015 年 1 月 3 日，窗口期是 1 天，那么，如果用户在2015年1月1日到2015年1月3日触发了步骤 A，并且在步骤 A 发生的 1 天内，依顺序依次触发了 B、C、D、E，则视作该用户完成了一次成功的漏斗转化。

在这个过程中，如果穿插了一些其它的步骤或者行为，例如在满足时间限制的情况下，用户的行为顺序是 A > X > B > X > C > D > X > E，X 代表任意一个事件，则该用户依然视作完成了一次成功的漏斗转化。

如果该用户在这个时间限制范围内，依次触发了 A > B > C > E，则该用户没有完成该漏斗的转化，并且会被记作步骤 C 的流失用户。

考虑一个更复杂的情况，如果一个用户在所选时段内有多个事件都符合某个转化步骤的定义，那么会优先选择更靠近最终转化目标的事件作为转化事件，并在第一次达到最终转化目标时停止转化的计算。假设一个漏斗的步骤定义是：访问首页、选择支付方式、支付成功，那么不同用户的行为序列及实际转化步骤（标红部分）见如下例子：

例 1：访问首页 -> 选择支付方式（支付宝） -> 选择支付方式（微信）-> 支付成功。

例 2：访问首页 -> 选择支付方式（支付宝） -> 访问首页 -> 选择支付方式（微信）-> 支付成功。

例 3：访问首页 -> 选择支付方式（支付宝） -> 访问首页 -> 选择支付方式（微信）-> 支付成功 -> 选择支付方式（微信）-> 支付成功。

**FAQ**

Q: 漏斗内的筛选条件和漏斗外的筛选条件的区别？

A: 漏斗内设置的筛选条件是根据设置的条件得到漏斗，漏斗外设置的筛选条件是根据得到的漏斗筛选出满足筛选条件的漏斗。一般业务上的使用场景是在漏斗内设置筛选条件，建议直接在漏斗内设置条件得到满足条件的漏斗。 比如漏斗选取的漏斗步骤为 A->B->C->D ，某用户甲的序列是 F->A2->A1->A1->B->C->A1->B，事件 A 包含了一个操作系统的属性，其中 A1 的操作系统为 Android， A2 的操作系统为 iOS。如果在漏斗内添加筛选条件操作系统为 iOS ，那么系统筛选到的序列为 A2->B->C，筛选的原则是沿着该用户的行为序列顺序查找，直到找到属性为 iOS 的 A 事件；如果在漏斗外添加筛选条件操作系统为 iOS 因为漏斗外筛选是在用户漏斗成功转换后的二次筛选，不加任何筛选条件时，该用户正常的漏斗转化为 A1->B->C,其中 A1 为第二个 A1 ，因此在漏斗外添加筛选条件时该用户筛选不出来

**Q: 分组的含义**

**A：**漏斗分析的分组，都是对完成转化/确认流失的用户的集合上进行分组。

漏斗分析的分组，包括二种不同的分组类型：

A：用户属性的分组：这个分组比较好理解，是在完成转化/确认流失的用户的集合，根据这个用户的属性，在来进行更进一步的分组。例如，我们添加的分组条件是“性别”，那么，就会分别对漏斗分析的结果按照“男”、“女”来进行分组；

B: 指定步骤的属性的分组：假设，我们选择了一个分组属性是步骤 2 的属性“支付方式”，这个筛选表示，在完成转化/确认流失的用户中，按照转化到步骤 2 时的“支付方式”的值来进行分组；如果有多次可能的转化，请参考基本计算规则中的说明；如果用户没有转化到步骤 2，则分到未知组。

**参考**

介绍及计算方法

<https://manual.sensorsdata.cn/sa/latest/guide_analytics_funnel-7540780.html>

实现分析

<https://tech.meituan.com/2018/03/20/user-funnel-analysis-design-build.html>

参考实现

<https://www.jianshu.com/p/4c86a2478cca>

-----------------------------------------------------------

留存分析

留存分析是一种用来分析用户参与情况/活跃程度的分析模型，考查进行初始行为后的用户中，有多少人会进行后续行为。这是衡量产品对用户价值高低的重要指标。留存分析中展示的数字代表独立用户数。表示在选定时间范围内进行了初始行为的用户，有多少人在随后的第 n 天/周/月进行了后续行为。

**基本计算规则**

假设定义的初始行为是 A 事件，后续行为是 B 事件，筛选时间段为 2015 年 1 月 1 日到 2015 年 1 月 8 日，注意这个时间范围是事件 A 发生的时间范围，事件 B 发生的时间范围是 2015 年 1 月 1日到 1 月 15 日（1 月 8 日加上 7 天）

**未设置关联属性**

下表为某用户2015 年 1 月 1 日到 2015 年 1 月 8 日的真实行为序列和纳入计算的行为序列（保留用户当日首个初始行为事件或后续行为事件）。表格中，字母 A 和 B 为事件，数字 1，2，3 为该用户某个属性的属性值。

| 日期 | 真实行为序列 | 纳入计算的行为序列 |
| --- | --- | --- |
| 01-01 | A1,A2,A2 | A1 |
| 01-02 | B2,B1,B1 | B2 |
| 01-03 | A3,A1,A1 | A3 |
| 01-04 | A1,A3,A3 | A1 |
| 01-05 | A1,A3,A3 | A1 |
| 01-06 | B1,B2,B1 | B1 |
| 01-07 | A1,A2 | A1 |
| 01-08 | B2,B1,B3 | B2 |

1.不加分组，如果指定初始行为日期为 2015 年 1 月 1 日，则该用户分别是第 1 天，第 5 天，第 7 天的留存用户。

2.按初始行为事件 A 的属性分组，如果用户完成事件 A 的属性值各不相同，该用户只会被归到 1 月 1 日 发生的首个 A 事件的属性值 1 中。去重后，该用户分别是属性值 1 的第 1 天，第 2 天，第 3 天，第 4 天，第 5 天，第 7 天的留存用户。

3.按后续行为事件 B 的属性分组，如果用户完成事件 B 的属性值各不相同，该用户只会被归到 1 月 2 日 发生的首个 B 事件的属性值 2 中。去重后，该用户分别是属性值 2 的第 1 天，第 2 天，第 3 天，第 4 天，第 5 天，第 7 天的留存用户。

1. 按用户属性分组，比如按性别分组，若用户为女性，则该用户分别是属性值女性的第 1 天，第 2 天，第 3 天，第 4 天，第 5 天，第 7 天的留存用户。

**设置关联属性**

下表为某用户 2015 年 1 月 1 日到 2015 年 1 月 8 日的真实行为序列和纳入计算的行为序列（同一天内初始行为事件或后续行为事件不同属性值各保留一个，且保留首个）。表格中，字母 A 和 B 为事件，数字 1，2，3 为该用户某个属性的属性值。

| 日期 | 真实行为序列 | 纳入计算的行为序列（属性值1） | 纳入计算的行为序列（属性值 2） |
| --- | --- | --- | --- |
| 01-01 | A1,A2,A2 | A1 | A2 |
| 01-02 | B2,B1,B1 | B1 | B2 |
| 01-03 | A3,A1,A1 | A1 |  |
| 01-04 | A1,A3,A3 | A1 |  |
| 01-05 | A1,A3,A3 | A1 |  |
| 01-06 | B1,B2,B1 | B1 | B2 |
| 01-07 | A1,A2 | A1 | A2 |
| 01-08 | B2,B1,B3 | B1 | B2 |

按初始行为事件 A 的属性分组，如果按属性值 1 分组，则关联属性值为 1 的初始行为事件或后续行为事件才会纳入计算，该用户分别是属性值 1 的第 1 天，第 2 天，第 3 天，第 4 天，第 5 天，第 7 天的留存用户。如果按属性值 2 分组，则关联属性值为 2 的初始行为事件或后续行为事件才会纳入计算，该用户分别是属性值 2 的第 1 天，第 5 天，第 7 天的留存用户。同理，该用户分别是属性值 3 的第 3 天，第 4 天，第 5 天的留存用户。

**FAQ**

Q: 筛选条件的含义？

A: 留存分析的筛选提供了两种不同的筛选类型。

用户属性上的筛选：例如，我们添加的筛选条件是“性别”为“男”，则只有属性中“性别”为“男”的用户，才满足这个筛选条件，并且出现在筛选后的留存分析结果中；

事件属性的筛选：和漏斗的触发限制条件含义相同，指定事件满足指定属性的过滤。

Q: 分组的含义？

A: 留存分析提供了两种不同的分组类型。我们以一个初始行为是 A，后续行为是 B，时间范围是 2015 年 1 月 1 日到 1 月 8 日的 7 天留存来进行详细说明：

用户属性上的分组：根据用户属性来进行更进一步的分组。例如我们添加的分组条件是“性别”，那么，就会分别对留存分析的结果按照“男”、“女”来进行分组；

事件属性的分组：例如，我们选择的分组设置是初始行为的属性“屏幕高度”，则这个分组表示，在 2015 年 1 月 1 日到 1 月 15 日这个时间范围内，按初始行为的“屏幕高度”这个属性的值来对他们进行分组；下面是几个具体的例子的描述:

某个用户在这个时间段内的行为序列是 A、B、C、A、B，第一次出现的 A 的“屏幕高度”值为“320”，第二次出现的 A 的“屏幕高度”值为“1080”，因为按照首次出现的 A 事件的“屏幕高度”来分组，所以这个用户会被划分到“320”这个分组的统计结果中;

某个用户在这个时间段内的行为序列是 A、A，这个用户在初始行为 A 事件后没有后续行为。第一次出现的 A 的“屏幕高度”值为“1080”，第二次出现的 A 的“屏幕高度”值为“320”，因为按照首次出现的 A 事件的“屏幕高度”来分组，所以这个用户会被划分到“1080”这个分组的统计结果中

**参考：**

介绍及计算方法<https://manual.sensorsdata.cn/sa/latest/guide_analytics_retention-1573691.html>

实现思路

<http://www.haoblogs.cn/blog/front/toDetail?id=34>

参考实现

<https://zhuanlan.zhihu.com/p/37907177?utm_source=wechat_session>

--------------------------------------------------------------------------------------------------

分布分析

分布分析用来分析某个事件指标的用户分布情况。比如，查看订单金额在 100 元以下、100 元至 200 元、200 元以上三个区间的用户分布情况。

**基本计算规则**

分布分析有三种统计方法，按时间段统计、按次数统计和按事件属性的统计指标。选择次数和事件的指标时可以自定义区间。

**按时间段统计**

统计用户在一天/周/月中，有多少个自然时间段（小时/天）进行了某项操作。

以『一天内』进行『点击广告』的『小时数』分析为例来说明。如果某用户在 15:00 到 16:00 间进行了『点击广告』3 次，17:00 到 18:00 间进行了『点击广告』1 次，则统计为『一天至少在 2 个时段进行点击广告』, 用户在某小时内进行『点击广告』一次或多次，都记 1 次。

对于查看用户一周/一月内某事件的天数，则是统计用户在一周/月中触发过某事件的天数。

**按次数统计**

统计用户在一天/周/月中，进行某项操作的次数，发生一次就记录一次。

**按事件属性的统计指标统计**

统计用户在一天/周/月中，发生事件的某属性的统计指标值。属性的统计指标与事件分析一致，有总和、均值、最大值、最小值、去重数。

**参考**

介绍及计算方法

<https://manual.sensorsdata.cn/sa/latest/guide_analytics_addiction-1573694.html>

----------------------------------------------------------------------------------

间隔分析

产品，运营，市场等人员的日常工作都需要观察某某业务的转化情况。如何衡量转化，除了用漏斗看转化率，还需要看转化时长的分布情况，间隔分析即是解决这类问题和需求的。通过计算用户行为序列中两个事件的时间间隔，得到业务转化环节的转化时长分布。

**基本计算规则**

假设用户在过去某个时间段内行为序列是：A → C → A → B → B → A → B → B → A → A → C → D → A → B

此时我们分析用户做了 A 事件和 B 事件的时间间隔，会按如下的计算规则：

在发生 A 事件后，找到离 A 事件 最近的 B 事件，即为第一个间隔。从间隔向后继续找 A 和 B 的配对，间隔与间隔不交叉，依次类推；

选择聚合时间单位，按天、周、月聚合，则会限定配对的 A 事件 和 B 事件发生在同一天、周、月；例如：如果按天聚合，用户发生A 事件 的时间是 23:50 ，发生 B 事件的时间是次日 00:10 。这两个事件无法完成配对。

此时如果不考虑聚合时间单位，则间隔配对结果是：A → C → A → B → B → A → B → B → A → A → C → D → A → B

**分析结果中各指标的计算规则**：

间隔数：在选定时间范围内，按照间隔分析计算规则完成间隔转化的配对数（下载数据中的字段）。

转化用户数：共有多少人在选定的时间内完成了间隔转化，可能一个人会完成多次间隔转化。

人均间隔数：间隔数 / 转化用户数（下载数据中的字段）。

间隔转化时长：在选定时间范围内，按照间隔分析计算规则，计算完成间隔转化的配对之后，统计的每个配对的间隔转化时长。

最大值：间隔转化时长的最大值。

最小值：间隔转化时长的最小值。

中位数：将间隔转化时长按从大到小排期，取中间值。

上四分位：将间隔转化时长按从大到小排期，取 1/4 处的值。

下四分位：将间隔转化时长按从大到小排期，取 3/4 处的值。

平均值：间隔转化时长的总和 / 转化用户数。

人均转化时间：每个人的平均间隔时长总和 / 转化用户数。

**参考**

介绍及计算方法https://manual.sensorsdata.cn/sa/latest/guide\_analytics\_attribution-1573697.html

------------------------------------------------------------------------------------------------

事件分析

事件，是追踪或记录的用户行为或业务过程。举例来说，一个电商产品可能包含如下事件：用户注册、浏览商品、添加购物车、支付订单等。

事件分析，是指基于事件的指标统计、属性分组、条件筛选等功能的查询分析。

**参考**

介绍及计算方法https://manual.sensorsdata.cn/sa/latest/guide\_analytics\_event-1573682.html