**DMP已有原始日志**

[平台维表 2](#_Toc355781895)

[1 亿起发 2](#_Toc355781896)

[1.1 广告点击 2](#_Toc355781897)

[1.2 订单 3](#_Toc355781898)

[2 购物客 4](#_Toc355781899)

[2.1 搜索 4](#_Toc355781900)

[3 亿告 5](#_Toc355781901)

[3.1 广告PV 5](#_Toc355781902)

[3.2 广告点击 5](#_Toc355781903)

[4 RTB 7](#_Toc355781904)

[4.1 请求 7](#_Toc355781905)

[4.2 竞价详情 7](#_Toc355781906)

[4.3 竞价成功及点击 8](#_Toc355781907)

[5 易博 9](#_Toc355781908)

[5.1 浏览 9](#_Toc355781909)

[5.2 商品展示 10](#_Toc355781910)

[5.3 事件 11](#_Toc355781911)

[5.4 订单 12](#_Toc355781912)

[6 Cookie 13](#_Toc355781913)

[6.1 易博、亿告、亿起发 13](#_Toc355781914)

|  |  |
| --- | --- |
| 平台维表 表1 | |
| 平台编号 | 平台名称 |
| 1 | 亿起发 |
| 2 | 亿告 |
| 3 | 易博 |
| 4 | 购物客 |
|  |  |
| 5010 | google |
| 5020 | tankx |
| 5030 | tencent |
| 50 | 广告交换 |
| 40 | 广告网络 |
| 30 | 社区 |
| 20 | 大媒体 |

# 亿起发

## 广告点击

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 数据 |
| 20130429 | 1 | 3362525 |
| 20130430 | 1 | 3221358 |
| 20130501 | 1 | 3311814 |
| 20130502 | 1 | 3484496 |
| 20130503 | 1 | 3523968 |
| 20130504 | 1 | 3560364 |
| 20130505 | 1 | 3578296 |

数据路径

Hive: s\_ad\_click\_log

hdfs:/data/stg/s\_ad\_click\_log/yyyyMMdd/1/

数据格式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 说明 | 举例 |
| click\_id | 点击流水ID | 必选 | 点击流水ID |  |
| pv\_id | 展现流水ID | 可选 | 展现流水ID |  |
| user\_ip | 点击用户的ip | 必选 | 点击用户的ip. 格式:255.255.255.255 | 125.118.79.238 |
| user\_id | 各平台内部唯一用户标志 | 必选 | 第三方平台唯一用户标志 |  |
| click time | 广告点击时间 | 必选 | click Time: 点击的时间,YYYYMMDDHH24MMSS | 20120522140909 |
| campaign\_id | 平台活动id | 必选 | 第三方平台下广告活动的ID |  |
| creative\_id | 平台创意id | 必选 | 第三方平台下广告创意的ID |  |
| ad\_modality | 广告形态 | 可选 | web广告/移动广告(web/moblie) |  |
| ad\_zone\_id | 广告位id | 必选 | 标识不同广告位置，便于统计不同广告位的效果 |  |
| user\_agent | 用户代理 | 可选 | 包含浏览器名称、版本，操作系统版本等信息 |  |

## 订单

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台编号 | 订单数 |
| 20130408 | 1 | 112498 |
| 20130409 | 1 | 122100 |
| 20130410 | 1 | 183517 |
| 20130411 | 1 | 118430 |
| 20130412 | 1 | 134594 |
| 20130413 | 1 | 104478 |
| 20130414 | 1 | 106825 |

说明20130415

数据路径

Hdfs: /data/stg/s\_order\_log/yyyyMMdd/1/

Hive: s\_order\_log

数据格式

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 英文名称 | 中午名称 | 可选项 | 格式 | 示例 | 说明 |
| order\_id | 订单唯一流水ID | 必选 |  |  |  |
| user\_id | 各平台内部唯一用户标志 | 必选 | 第三方平台唯一用户标志 |  | 第三方平台唯一用户标志 |
| order\_no | 订单号 | 必选 |  |  |  |
| product\_no | 商品编号 | 可选 |  |  | 第三方商品编号 |
| product\_name | 商品名称 | 必选 |  |  | 第三方商品名称 |
| product\_price | 商品单价 | 必选 |  | 70.5 | 单位为元 |
| product\_count | 商品数量 | 可选 |  | 2 | 默认为1 |
| product\_type\_id | 商品分类编号 | 可选 |  | 234 | 第三方商品分类编号 |
| product\_type\_name | 商品分类名称 | 可选 |  | 化妆品 |  |
| order\_price | 订单总价 | 必选 |  |  |  |
| campaign\_id | 平台活动id | 可选 |  |  |  |
| order\_time | 商品下单时间 | 必选 | YYYYMMDDHH24MMSS | 如:20120522140909 |  |
| domain | 点击域名 | 必选 |  |  | 消费域名投放需求新增 |

# 购物客

## 搜索

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 搜索数据 |
| 20130429 | 4 | 45306 |
| 20130430 | 4 | 45516 |
| 20130501 | 4 | 60669 |
| 20130502 | 4 | 57098 |
| 20130503 | 4 | 89907 |
| 20130504 | 4 | 104474 |
| 20130505 | 4 | 109799 |

数据路径

Hive: s\_search\_log

Hdfs:/data/stg/s\_search\_log/yyyyMMdd/4

数据格式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 说明 | 例子 |
| user\_id | 用户标识 | 必选 | 第三方平台唯一用户标志 |  |
| user\_ip | Ip地址 | 必选 | 浏览器客户端的IP地址 |  |
| timestamp | pv时间 | 必选 | 服务器时间 YYYYMMDDHH24MMSS | 20120522140909 |
| search\_word | 用户搜索词 | 必选 | 用户搜索词 |  |
| seller\_id | 商家id | 可选 | 商品所属商家id |  |
| catid | 商品类目id | 可选 | 商品所属类目id |  |
| high\_price | 商品价格上限 | 可选 | 商品价格范围 |  |
| low\_price | 商品价格下限 | 可选 | 商品价格范围 |  |

# 亿告

## 广告PV

数据规模

|  |  |
| --- | --- |
| **日期** | **数据** |
| 20130429 | 62129225 |
| 20130430 | 63008480 |
| 20130501 | 62982307 |
| 20130502 | 55604261 |
| 20130503 | 56104779 |
| 20130504 | 59887878 |
| 20130505 | 63235875 |

数据路径

|  |
| --- |
| Hive: s\_ad\_pv\_log |
| hdfs:/data/stg/s\_ad\_pv\_log/yyyyMMdd/2/ |

数据格式

|  |  |  |  |
| --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 说明 |
| pv\_id | 展示流水ID | 必选 | 第三方平台唯一展示表示 |
| user\_id | 各平台内部唯一用户标志 | 必选 | 第三方平台唯一用户标志 |
| user\_ip | Ip地址 | 可选 | 浏览器客户端的IP地址 |
| timestamp | pv时间 | 必选 | 服务器时间 YYYYMMDDHH24MMSS |
| page\_url | 当前页的地址 | 可选 | 浏览页面的地址 |
| referer\_url | 来源url | 可选 | 浏览页面的来源URL地址 |
| user\_agent | 用户代理 | 可选 | 包含浏览器名称、版本，操作系统版本等信息 |
| ad\_modality | 广告形态 | 可选 | web广告/移动广告(web/moblie) |
| ad\_zone\_id | 广告位id | 必选 | 标识某个广告位 |
| campaign\_id | 平台活动id | 必选 | 第三方平台下广告活动的ID |
| creative\_id | 平台创意ID | 必选 | 第三方平台下广告创意的ID |

## 广告点击

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 广告点击数据 |
| 20130429 | 2 | 12371 |
| 20130430 | 2 | 13432 |
| 20130501 | 2 | 13522 |
| 20130502 | 2 | 13579 |
| 20130503 | 2 | 16136 |
| 20130504 | 2 | 14195 |
| 20130505 | 2 | 12471 |

数据路径

Hive：s\_ad\_click\_log

Hdfs :/data/stg/s\_ad\_click\_log/yyyyMMdd/2/

数据格式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 说明 | 举例 |
| click\_id | 点击流水ID | 必选 | 点击流水ID |  |
| pv\_id | 展现流水ID | 可选 | 展现流水ID |  |
| user\_ip | 点击用户的ip | 必选 | 点击用户的ip. 格式:255.255.255.255 | 125.118.79.238 |
| user\_id | 各平台内部唯一用户标志 | 必选 | 第三方平台唯一用户标志 |  |
| click time | 广告点击时间 | 必选 | click Time: 点击的时间,YYYYMMDDHH24MMSS | 20120522140909 |
| campaign\_id | 平台活动id | 必选 | 第三方平台下广告活动的ID |  |
| creative\_id | 平台创意id | 必选 | 第三方平台下广告创意的ID |  |
| ad\_modality | 广告形态 | 可选 | web广告/移动广告(web/moblie) |  |
| ad\_zone\_id | 广告位id | 必选 | 标识不同广告位置，便于统计不同广告位的效果 |  |
| user\_agent | 用户代理 | 可选 | 包含浏览器名称、版本，操作系统版本等信息 |  |

# RTB

## 请求

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 数据 |
| 20130504 | 5010 | 582872198 |
| 20130504 | 5020 | 607856391 |
| 20130505 | 5010 | 583714105 |
| 20130505 | 5020 | 612594401 |

数据路径

Hive : s\_adx\_request\_log

Hdfs ：/data/stg/s\_adx\_request\_log/yyyyMMdd/

数据格式：

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 属性 | 含义 | 数据类型和长度 |
| Plat\_id | required | 供应商（比如google）的编号，对应请求的id | string |
| requestid | required | 谷歌唯一的请求id | 16字节 |
| 用户IP | optional | 访问媒体的用户ip地址 |  |
| google\_user\_id | optional | 用户cookie | string |
| URL | optional | 访问网址 |  |
| detected\_vertical | required | 媒体id | 多个用逗号隔开 |
|  | weight | 权重 |  |

## 竞价详情

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 数据 |
| 20130504 | 5010 | 14254752 |
| 20130504 | 5020 | 28257405 |
| 20130505 | 5010 | 3627777 |
| 20130505 | 5020 | 8274879 |

数据路径

Hive:s\_camp\_log

Hdfs : /data/stg/s\_camp\_log/yyyyMMdd/

数据格式

|  |  |  |
| --- | --- | --- |
| **字段名称** | **数据类型** | **说明** |
| 供应商id | string | 供应商（比如google）的编号 |
| 竞价的RequestID | string | 对于这次请求的标识，竞价的RequestID现在不是存原始值，而是存十六进制表示形式。总共32个十六进制字符(0-9, A-F). |
| 时间 | string | 本次请求的时间 |
| 竞价的UserId | string | 通过用户id来做关键信息，具体的用户属性信息，则需要根据用户id去查询相应的数据表 |
| 媒体类型id | string | 包括了媒体的类别编码，访问的URL，以及广告位尺寸，这是从竞价请求中能够得到的信息，但因为是可选项，不一定都带 |
| 媒体类型权重 | String | 谷歌可能会过来多个媒体类型id，每个对应一个权重 |
| 访问URL | string | landingpage的url |
| 广告位宽 | int | 广告位宽尺寸（mm） |
| 广告位高 | int | 广告位高尺寸（mm） |
| 创意类型 | string | 由于同用户信息的重复冗余，提供的信息不是很多，暂时将创意类型加入 |
| 竞价方式 | string | rt，bt,rel等, 分别用0、1、2表示 |
| 报价 | folate | 向Ad exchange的报价 |
| 活动id | String | 返回广告的活动（campaign）id |
| 创意id | String | 返回广告的创意（creative）id |
| 预测转化率 | float |  |
| 预测成功率 | float |  |
| 预测点击率 | float |  |
| 用户ip | string | 点分IP |
| 行政区域编码 | string | 广告交换平台给的中国行政区域编码，如google的CN-11,代表北京 |

## 竞价成功及点击

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 数据 |
| 20130504 | 5010 | 4324409 |
| 20130504 | 5020 | 6616132 |
| 20130505 | 5010 | 1190123 |
| 20130505 | 5020 | 2286983 |

数据路径

Hive: s\_camp\_succ\_log

Hdfs : /data/stg/s\_camp\_succ\_log/yyyyMMdd/

数据格式：

|  |  |  |
| --- | --- | --- |
| **字段名称** | **数据类型** | **说明** |
| 竞价的RequestID | 字符串 |  |
| 竞价的UserId | 字符串 | 即RTB接收到的用户唯一标识 |
| 广告平台标识 | 字符串 | AdEx/XXXX 此值由RTB定 |
| 计费类型 | 整型 | CPM/CPC (不区分大小写) |
| 动作类型 | 整型 | IMPRESSION -1/CLICK-2 (不区分大小写) |
| 竞价成功价格 | 浮点型 | 广告平台收取的价格(单位是0.0000001元) |
| 收取价格 | 浮点型 | 是DSP向广告主收取的价格，即1次展现或点击的应收客户的价格(单位是0.0000001元),预留字段当前为空 |
| 广告计划ID | 字符串 |  |
| 广告创意ID | 字符串 |  |
| 广告主ID | 字符串 |  |
| 项目ID | 字符串 |  |
| 竞价时间 | 时间 | 时间格式(YYYYMMDDHHmmss) |
| 广告浏览用户ID | 字符串 | 从当前域的cookie中得到的广告浏览用户唯一标识)(DMP\_USER\_ID) |
| 浏览用户IP | 字符串 |  |
| 浏览用户浏览器 | 字符串 |  |
| 广告位页面url | 字符串 |  |
| RTB出价 | 浮点型 | 是RTB向AdExchange的出价价格，即1次展现或点击向广告网络支付的价格(单位是0.0000001元) |
| CTR | 浮点型 | 收费类型为CPC时计费用 |
| 佣金比例 | 浮点型 |  |
| 税点比例 | 浮点型 |  |
| 流水时间 | 时间 | 时间格式(YYYYMMDDHHmmss) |
| 用户价值 | 浮点型 | 每次展现对广告主网站的价值，由RTB预算估值 |
| 利差分成 | 浮点型 | 我们赚取的**利润**分给广告主的比例 |

# 易博

## 浏览

浏览日志：  
1.含有易购的一些数据

2.浏览日志包含商品展示  
3.搜索广告，涉及百度（86110/小时）、搜狗（130/小时）。包含易购、中粮、高街、给力团等。

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 浏览数据 |
| 20130429 | 3 | 16590569 |
| 20130430 | 3 | 15116959 |
| 20130501 | 3 | 15052542 |
| 20130502 | 3 | 18912713 |
| 20130503 | 3 | 18828340 |
| 20130504 | 3 | 16654812 |
| 20130505 | 3 | 16117586 |

数据路径

Hive : s\_browse\_log

Hdfs : /data/stg/s\_browse\_log/yyyyMMdd/3

数据格式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 说明 | 例子 |
| user\_id | 用户标识 | 必选 | 第三方平台唯一用户标志 |  |
| user\_ip | Ip地址 | 必选 | 浏览器客户端的IP地址 |  |
| timestamp | pv时间 | 必选 | 服务器时间 YYYYMMDDHH24MMSS | 20120522140909 |
| page\_url | 当前页的地址 | 必选 | 浏览页面的地址 |  |
| referer\_url | 来源url | 可选 | 浏览页面的来源URL地址 |  |
| keywords | 关键字 | 可选 | 浏览当前页的关键字 |  |
| page\_title | 标题 | 可选 | 浏览当前页的标题 |  |
| description | 简介 | 可选 | 浏览当前页的简介 |  |
| user\_agent | 用户代理 | 可选 | 包含浏览器名称、版本，操作系统版本等信息 |  |
| stay\_time | 停留时常 | 可选 |  |  |
| rule\_id | 规则id | 可选 | 活动的规则id |  |

## 商品展示

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台编号 | 数据 |
| 20130429 | 3 | 1278499 |
| 20130430 | 3 | 1141911 |
| 20130501 | 3 | 1140978 |
| 20130502 | 3 | 1498485 |
| 20130503 | 3 | 1521083 |
| 20130504 | 3 | 1275068 |
| 20130505 | 3 | 1213323 |

数据路径

Hive : s\_browse\_prod\_log

Hdfs: /data/stg/s\_browse\_prod\_log/yyyyMMdd/3

数据格式

|  |  |  |  |
| --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 例子 |
| plat\_user\_id | 用户标识 | 必选 |  |
| ip | Ip地址 | 必选 |  |
| oper\_time | pv时间 | 必选 | 20120522140909 |
| page\_url | 当前页的地址 | 必选 |  |
| refer\_url | 来源url | 可选 |  |
| prod\_no | 商品编号 | 可选 |  |
| prod\_name | 商品名称 | 可选 |  |
| prod\_price | 商品价格 | 可选 |  |
| orig\_prod\_type | 原始商品分类编号 | 可选 |  |
| prod\_type\_name | 商品分类名称 | 可选 |  |
| coll\_type | 商品收集分类 | 可选 |  |

## 事件

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 数据 |
| 20130429 | 3 | 84421 |
| 20130430 | 3 | 78082 |
| 20130501 | 3 | 76796 |
| 20130502 | 3 | 95338 |
| 20130503 | 3 | 92639 |
| 20130504 | 3 | 85142 |
| 20130505 | 3 | 86752 |

数据路径

Hive：s\_event\_log

Hdfs :/data/stg/s\_event\_log/yyyyMMdd/

数据格式

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 英文名称 | 中午名称 | 可选项 | 格式 | 示例 | 说明 |
| user\_id | 各平台内部唯一用户标志 | 必选 | 第三方平台唯一用户标志 |  | 第三方平台唯一用户标志 |
| user\_ip | 客户端IP | 必选 |  | 125.118.79.238 |  |
| url | 事件url | 必选 |  |  |  |
| event\_id | 事件id | 必选 |  |  | 注册事件,登录事件等，参考事件属性接口 |
| timestamp | 事件时间 | 必选 | YYYYMMDDHH24MMSS | 如:20120522140909 |  |
| user\_agent | 用户代理 | 可选 |  |  | 包含浏览器名称、版本，操作系统版本等信息 |

## 订单

数据规模

|  |  |  |
| --- | --- | --- |
| 日期 | 平台标识 | 订单数据 |
| 20130429 | 3 | 48746 |
| 20130430 | 3 | 50541 |
| 20130501 | 3 | 49457 |
| 20130502 | 3 | 57553 |
| 20130503 | 3 | 60642 |
| 20130504 | 3 | 47460 |
| 20130505 | 3 | 47088 |

数据路径

Hive: s\_order\_log

Hdfs:/data/stg/s\_order\_log/yyyyMMdd/3/

数据格式

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 英文名称 | 中午名称 | 可选项 | 格式 | 示例 | 说明 |
| order\_id | 订单唯一流水ID | 必选 |  |  |  |
| user\_id | 各平台内部唯一用户标志 | 必选 | 第三方平台唯一用户标志 |  | 第三方平台唯一用户标志 |
| order\_no | 订单号 | 必选 |  |  |  |
| product\_no | 商品编号 | 可选 |  |  | 第三方商品编号 |
| product\_name | 商品名称 | 必选 |  |  | 第三方商品名称 |
| product\_price | 商品单价 | 必选 |  | 70.5 | 单位为元 |
| product\_count | 商品数量 | 可选 |  | 2 | 默认为1 |
| product\_type\_id | 商品分类编号 | 可选 |  | 234 | 第三方商品分类编号 |
| product\_type\_name | 商品分类名称 | 可选 |  | 化妆品 |  |
| order\_price | 订单总价 | 必选 |  |  |  |
| campaign\_id | 平台活动id | 可选 |  |  |  |
| order\_time | 商品下单时间 | 必选 | YYYYMMDDHH24MMSS | 如:20120522140909 |  |
| domain | 点击域名 | 必选 |  |  | 消费域名投放需求新增 |

# Cookie

## 易博、亿告、亿起发

数据规模

|  |  |
| --- | --- |
| 日期 | 新增 |
| 20130429 | 27614908 |
| 20130430 | 27410239 |
| 20130501 | 28842698 |
| 20130502 | 26784003 |
| 20130503 | 27105873 |
| 20130504 | 26122804 |
| 20130505 | 25263897 |

数据路径

Hive: s\_cookie\_log

Hdfs :/data/stg/s\_cookie\_log/yyyyMMdd/

数据格式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 英文名称 | 中文名称 | 可选项 | 格式 | 示例 |
| user\_id | dmpid | 必选 |  | dmp cookieid |
| orig\_plat\_type | 平台id | 必选 |  | 请参照平台维表 |
| plat\_user\_id | 各平台内部唯一用户标志 | 必选 |  | 第三方平台唯一用户标志 |
| create\_time | 创建时间 | 必选 | YYYYMMDDHH24MMSS | 如:20120522140909 |