**大数据项目一**

**项目背景**

电商日志分析平台

--pageview（pv）：网站页面的访问量;当用户点击某个网站的页面的时候，pv值加1。

借助站长工具（http://seo.chinaz.com/www.dangdang.com）查看网站页面的日均访问量（pv）。

--以当当网站为例：

假设每条pv会产生1k的数据。那么对于日均pv（100万）就会产生大约1G的数据。

这只是涉及到网站页面的浏览产生的数据，对于其他的用户行为（用户的下单操作）产生的日志信息产生的数据会更多。

--电商日志分析平台（公共的、成熟的分析平台）：百度统计、Google分析、易分析。

tongji.baidu.com



我们自己的系统完全可以整合到百度统计和Google分析。方法就是注册一个账号，然后下载其js代码，嵌套到自己的业务系统中。这样我们在登录后就能看到自己系统的分析情况。

--对于这些成熟的网站统计，我们为什么还要自己去设计日志分析平台呢？？

自己设计的可以实现个性化，更重要的是为了数据的安全考虑。刚起步的小公司初期可能会使用成熟的分析平台。

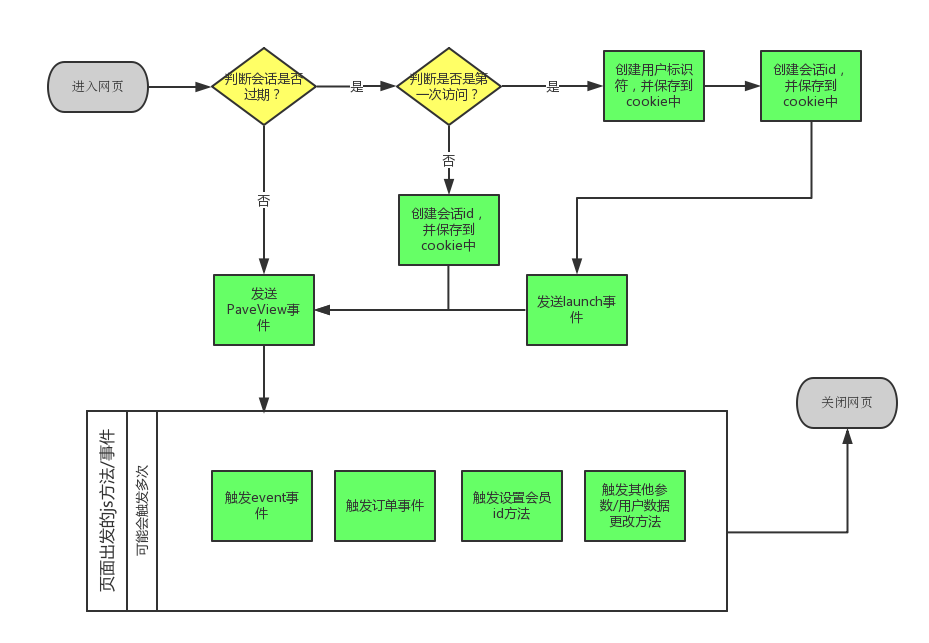
# 概述

该文档的主要作用是为了开发人员参考可以参考本文档，了解js sdk的集成方式以及提供的各种不同的api。

注意：不采用ip来标示用户的唯一性，我们通过在cookie中填充一个uuid来标示用户的唯一性。

# Js sdk执行工作流

在我们的js sdk中按照收集数据的不同分为不同的事件，比如pageview事件等。Js sdk的执行流程如下：



# 分析

1. **PC端事件分析**

针对我们最终的不同分析模块，我们需要不同的数据，接下来分别从各个模块分析，每个模块需要的数据。用户基本信息就是用户的浏览行为信息分析，也就是我们只需要pageview事件就可以了；浏览器信息分析以及地域信息分析其实就是在用户基本信息分析的基础上添加浏览器和地域这个维度信息，其中浏览器信息我们可以通过浏览器的window.navigator.userAgent来进行分析，地域信息可以通过nginx服务器来收集用户的ip地址来进行分析，也就是说pageview事件也可以满足这两个模块的分析。外链数据分析以及用户浏览深度分析我们可以在pageview事件中添加访问页面的当前url和前一个页面的url来进行处理分析，也就是说pageview事件也可以满足这两个模块的分析。订单信息分析要求pc端发送一个订单产生的事件，那么对应这个模块的分析，我们需要一个新的事件chargeRequest。对于事件分析我们也需要一个pc端发送一个新的事件数据，我们可以定义为event。除此之外，我们还需要设置一个launch事件来记录新用户的访问。

Pc端的各种不同事件发送的数据url格式如下,其中url中后面的参数就是我们收集到的数据：http://bjsxt.com/bjsxt.gif?requestdata

|  |  |
| --- | --- |
| **最终分析模块** | **PC端js sdk事件** |
| 用户基本信息分析 | pageview事件 |
| 浏览器信息分析 |
| 地域信息分析 |
| 外链数据分析 |
| 用户浏览深度分析 |
| 订单信息分析 | chargeRequest事件 |
| 事件分析 | event事件 |
|  | launch事件 |

1. **Launch事件**

当用户第一次访问网站的时候触发该事件，不提供对外调用的接口，只实现该事件的数据收集。

|  |  |
| --- | --- |
| 方法名称 | - |
| 发送的数据 | en=e\_l&ver=1&pl=website&sdk=js&u\_ud=7096A1DE-79E0-45BF-97F2-C64CEDE27579&u\_sd=4BD5846B-2162-41B4-8255-51A0E59E594F&c\_time=1506261590411&l=zh-CN&b\_iev=Mozilla%2F5.0%20(Windows%20NT%206.1%3B%20Win64%3B%20x64%3B%20rv%3A51.0)%20Gecko%2F20100101%20Firefox%2F51.0&b\_rst=1366\*768 |

1. **Pageview事件**

当用户访问页面/刷新页面的时候触发该事件。该事件会自动调用，也可以让程序员手动调用。

|  |  |
| --- | --- |
| 方法名称 | onPageView |
| 发送的数据 | en=e\_pv&p\_url=http%3A%2F%2Flocalhost%3A8080%2FBIG\_DATA\_LOG2%2Fdemo.jsp&tt=%E6%B5%8B%E8%AF%95%E9%A1%B5%E9%9D%A21&ver=1&pl=website&sdk=js&u\_ud=7096A1DE-79E0-45BF-97F2-C64CEDE27579&u\_sd=D6F4FA18-A13F-4850-91F4-D9DCF7B751FD&c\_time=1506262358144&l=zh-CN&b\_iev=Mozilla%2F5.0%20(Windows%20NT%206.1%3B%20Win64%3B%20x64%3B%20rv%3A51.0)%20Gecko%2F20100101%20Firefox%2F51.0&b\_rst=1366\*768 |

1. **chargeRequest事件**

当用户下订单的时候触发该事件，该事件需要程序主动调用。

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名称 | | onChargeRequest | |
| 发送的数据 | | u\_sd=8E9559B3-DA35-44E1-AC98-85EB37D1F263&c\_time=1449139048231&oid=orderid123&on=%E4%BA%A7%E5%93%81%E5%90%8D%E7%A7%B0&cua=1000&cut=%E4%BA%BA%E6%B0%91%E5%B8%81&pt=%E6%B7%98%E5%AE%9D&ver=1&en=e\_crt&pl=website&sdk=js&b\_rst=1920\*1080&u\_ud=12BF4079-223E-4A57-AC60-C1A04D8F7A2F&b\_iev=Mozilla%2F5.0%20(Windows%20NT%206.1%3B%20WOW64)%20AppleWebKit%2F537.1%20(KHTML%2C%20like%20Gecko)%20Chrome%2F21.0.1180.77%20Safari%2F537.1&l=zh-CN | |
| 参数 | 类型 | 是否必填 | 描述 |
| orderId | string | 是 | 订单id |
| orderName | String | 是 | 产品购买描述名称 |
| currencyAmount | double | 是 | 订单价格 |
| currencyType | String | 是 | 货币类型 |
| paymentType | String | 是 | 支付方式 |

1. **Event事件**

当访客/用户触发业务定义的事件后，前端程序调用该方法。

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名称 | | onEventDuration | |
| 发送的数据 | | ca=%E7%B1%BB%E5%9E%8B&ac=%E5%8A%A8%E4%BD%9C&c\_time=1449139512665&u\_sd=8E9559B3-DA35-44E1-AC98-85EB37D1F263&kv\_p\_url=http%3A%2F%2Fwwwbjsxt..com%3A8080%2Fvst\_track%2Findex.html&kv\_%E5%B1%9E%E6%80%A7key=%E5%B1%9E%E6%80%A7value&du=1000&ver=1&en=e\_e&pl=website&sdk=js&b\_rst=1920\*1080&u\_ud=12BF4079-223E-4A57-AC60-C1A04D8F7A2F&b\_iev=Mozilla%2F5.0%20(Windows%20NT%206.1%3B%20WOW64)%20AppleWebKit%2F537.1%20(KHTML%2C%20like%20Gecko)%20Chrome%2F21.0.1180.77%20Safari%2F537.1&l=zh-CN | |
| 参数 | 类型 | 是否必填 | 描述 |
| category | string | 是 | 自定义事件名称 |
| action | String | 是 | 自定义事件动作 |
| map | map | 否 | 其他参数 |
| duration | long | 否 | 事件持续时间 |

1. **其他api方法**

在这里只介绍设置会员id的方法，其他的辅助方法到时候编写js的时候再介绍。

**a)设置会员id**

|  |  |  |  |
| --- | --- | --- | --- |
| 方法名称 | | setMemberId | |
| 方法描述 | | 该方法的主要作用是设置会员id，当会员登录后，请调用该方法进行会员id的设置。 | |
| 参数 | 类型 | 是否必填 | 描述 |
| memberid | string | 是 | 会员id |

1. **数据参数说明**

在各个不同事件中收集不同的数据发送到nginx服务器，但是实际上这些收集到的数据还是有一些共性的。下面将所用可能用到的参数描述如下：

|  |  |  |
| --- | --- | --- |
| 参数名称 | 类型 | 描述 |
| en | string | 事件名称, eg: e\_pv |
| ver | string | 版本号, eg: 0.0.1 |
| pl | string | 平台, eg: website |
| sdk | string | Sdk类型, eg: js |
| b\_rst | string | 浏览器分辨率，eg: 1800\*678 |
| b\_iev | string | 浏览器信息useragent |
| u\_ud | string | 用户/访客唯一标识符 |
| l | string | 客户端语言 |
| u\_mid | string | 会员id，和业务系统一致 |
| u\_sd | string | 会话id |
| c\_time | string | 客户端时间 |
| p\_url | string | 当前页面的url |
| p\_ref | string | 上一个页面的url |
| tt | string | 当前页面的标题 |
| ca | string | Event事件的Category名称 |
| ac | string | Event事件的action名称 |
| kv\_\* | string | Event事件的自定义属性 |
| du | string | Event事件的持续时间 |
| oid | string | 订单id |
| on | string | 订单名称 |
| cua | string | 支付金额 |
| cut | string | 支付货币类型 |
| pt | string | 支付方式 |