

XX 电商用户行为分析系统（一）

目录

课程目标.....	2
1、项目背景.....	2
1.1、概述.....	2
1.2、核心指标概念详解.....	2
2、点击流数据模型.....	3
2.1、日志.....	3
2.2、点击流概念.....	4
2.3、点击流数据模型.....	4
2.4、网站流量数据分析的意义.....	5
2.5、如何进行网站流量分析.....	5
2.5.1、网站流量质量分析.....	6
2.5.2、网站流量多维度细分.....	6
2.5.3、网站内容及导航分析.....	7
2.5.4、网站转化及漏斗分析.....	8
3、流量分析常见指标.....	9
3.1、基础分析.....	9
3.2、来源分析.....	10
3.3、受访分析.....	12
3.4、访客分析.....	13
3.5、转化路径分析.....	15
3.6、用户分析模块.....	16
3.7、浏览器分析模块.....	17
3.8、地域分析模块.....	17
3.9、用户浏览深度分析模块.....	18
3.10、订单分析模块.....	18
3.11、热力图分析.....	18
3.12、投入产出分析（ROI）.....	19
4、项目技术架构.....	19
4.1、数据处理流程.....	19
4.2、技术选型.....	20
4.2.1、数据收集层.....	20
4.2.2、数据分析层.....	20
4.2.3、数据展示层.....	20
4.3、项目结构.....	21
4.4、总体架构分析.....	21
4.5、数据展现.....	22
5、集群开发环境准备.....	22
5.1、集群规划.....	22
5.2、环境搭建.....	23

课程目标

理解网站点击流数据分析的**业务背景**

理解网站点击流数据分析中**常用分析指标**的业务含义

掌握网站点击流数据分析系统的**技术架构**

掌握网站点击流数据分析系统中各环节的**技术实现**

能独立设计完成一个初步的**网站点击流数据分析系统**

网站点击流数据分析，业务知识，推荐书籍：

《网站分析实战——如何以数据驱动决策，提升网站价值》王彦平，吴盛锋编著

1、项目背景

1.1、概述

当今社会有很多的电商网站，这就存在着一些竞争关系，为了更好的设计一个网站，让一个电商网站浏览的人数更多，从而增加点击量和订阅量的，这样就需要我们对这个电商网站进行分析和数据挖掘，我们可以**根据每天浏览某电商网站的人数和访客量来判断一个网站的好坏和受欢迎程度，同时也可以根据外链的跳转率和访客或会员所用的浏览器等工具的分析来进行精准的广告推广，我们也可以根据地区的点击量和访客或是会员访问的时间的分析来进行合理的商品推广，精准的推荐**等操作。同时每一个电商网站，可以根据这个网站的支付订单数以及成功支付订单数来进行业务的分析，这些对于提高一个网站的点击量、浏览量、以及成功支付订单量都是必不可少的。

即指用户访问网站时的所有访问、浏览、点击行为数据。比如点击了哪一个链接，在哪个网页停留时间最多，采用了哪个搜索项、总体会话时间等。而所有这些信息都可被保存在网站日志中。通过分析这些数据，可以获知许多对网站运营至关重要的信息。采集的数据越全面，分析就能越精准。

总之，一个电商网站就应该设计出一款产品能让用户的体验好，能让用户精准的寻找想要购买的商品，能提高用户的转化率，能提高广告的转化率。

1.2、核心指标概念详解

PV

UV

独立 IP

访问深度

转化率

跳出率
ROI
RFM
订单数
订单总金额
客单价
.....

2、点击流数据模型

2.1、日志

日志的生成渠道：

- 1、是网站的 web 服务器软件（apache、nginx、tomcat）所记录的 web 访问日志；
- 2、是通过在页面嵌入自定义的 JS 代码来获取用户的所有访问行为（比如鼠标悬停的位置，点击的页面组件等），然后通过 AJAX 请求到后台记录日志；这种方式所能采集的信息最全面；
- 3、通过在页面上埋点 1 像素的图片，将相关页面访问信息请求到后台记录日志；

日志数据内容详述：

在实际操作中，有以下几个方面的数据可以被采集：

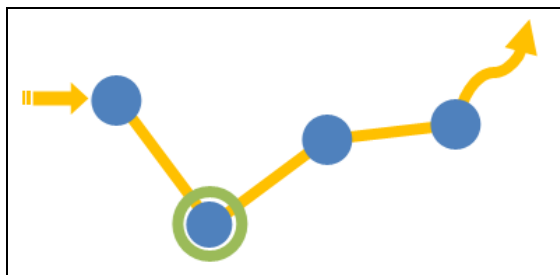
- 1、访客的系统属性特征。比如所采用的操作系统、浏览器、域名和访问速度等。
- 2、访问特征。包括停留时间、点击的 URL、所点击的“页面标签<a>”及标签的一些属性（比如业务 entity<商品、电影、歌曲、小说名称>的名称）等。
- 3、来源特征。包括来访 URL，来访 IP 等。
- 4、产品特征。包括所访问的产品编号、产品类别、产品颜色、产品价格、产品利润、产品数量和特价等级等。

以某电商日志数据为例子：

```
GET
/log.gif?t=item.010001&m=UA-J2011-1&pin=-
&uid=1679790178&sid=1679790178|12&v=je=1$sc=24-bit$sr=1600x900$ul=zh-
cn$cs=GBK$dt=xxxx$hn=item.jd.com$fl=16.0r0$os=win$br=chrome$bv=39.0.2171.95$wb=14
37269412$xb=1449548587$yb=1456186252$zb=12$cb=4$usc=direct$ucp=-
$umd=none$uct=-$ct=1456186505411$lt=0$tad=-
$sku=1326523$cid1=1316$cid2=1384$cid3=1405$brand=20583$pinid=-
&ref=&rm=1456186505411 HTTP/1.1
```

2.2、点击流概念

点击流这个概念更注重用户浏览网站的整个流程，网站日志中记录的用户点击就像是图上的“点”，而点击流更像是将这些“点”串起来形成的“线”。也可以把“点”认为是网站的 Page，而“线”则是访问网站的 Session。所以点击流数据是由网站日志中整理得到的，它可以比网站日志包含更多的信息，从而使基于点击流数据统计得到的结果更加丰富和高效。



2.3、点击流数据模型

点击流数据在具体操作上是由散点状的点击日志数据梳理所得，从而，点击数据在数据建模时应该存在两张模型表（pageviews 和 visits）：

1、用于生成点击流的原始访问日志表

时间戳	IP 地址	请求 URL	Referral	响应码	流量	br 信息
2012-01-01 12:31:12	101.0.0.1	/a/...	somesite.com			
2012-01-01 12:31:16	201.0.0.2	/a/...	aura.cn			
2012-01-01 12:33:06	101.0.0.2	/b/...	baidu.com			
2012-01-01 15:16:39	234.0.0.3	/c/...	google.com			
2012-01-01 15:17:11	101.0.0.1	/d/...	/c/...			
2012-01-01 15:19:23	101.0.0.1	/e/...	/d/...			

2、页面点击流模型 pageviews 表

Session	IP 地址	时间	访问页面 URL	停留时长	第几步
S001	101.0.0.1	2012-01-01 12:31:12	/a/....	30	1
S002	201.0.0.2	2012-01-01 12:31:16	/a/....	10	1
S002	201.0.0.2	2012-01-01 12:33:06	/b/....	110	2
S002	201.0.0.2	2012-01-01 12:35:06	/e/....	30	3
S003	201.0.0.2	2012-01-01 15:35:06	/a/....	20	1

3、点击流模型 visits 表(按 session 聚集的页面访问信息)

Session	起始时间	结束时间	进入页面	离开页面	访问页面数	IP	cookie	referral
S001	2012-01-01 12:31:12	2012-01-01 12:31:12	/a/...	/a/...	1	101.0.0.1	User01	somesite.com
S002	2012-01-01	2012-01-01	/a/...	/e/...	3	201.0.0.2	User02	-

	12:31:16	12:35:06						
S003	2012-01-01 12:35:42	2012-01-01 12:35:42	/c/...	/c/...	1	234.0.0.3	User03	baidu.co m
S003	2012-01-01 15:16:39	2012-01-01 15:19:23	/c/...	/e/...	3	101.0.0.1	User01	google.co m
.....

这就是点击流模型。当 WEB 日志转化成点击流数据的时候，很多网站分析度量的计算变得简单了，这就是点击流的“魔力”所在。**基于点击流数据我们可以统计出许多常见的网站分析度量指标**

2.4、网站流量数据分析的意义

网站流量统计分析，可以帮助网站管理员、运营人员、推广人员等实时获取网站流量信息，并从流量来源、网站内容、网站访客特性等多方面提供网站分析的数据依据。从而帮助提高网站流量，提升网站用户体验，让访客更多的沉淀下来变成会员或客户，通过更少的投入获取最大化的收入。

如下表：

网站的眼睛	网站的神经	网站的大脑
访问者来自哪里？ 访问者在寻找什么？ 哪些页面最受欢迎？ 访问者从哪里进入？ 访问者从哪里跳出？	网页布局合理吗？ 网站导航清晰吗？ 哪些功能存在问题 网站内容有效吗 转化路径靠谱吗？	如何分解目标？ 如何分配广告预算？ 如何衡量产品表现？ 哪些产品需要优化？ 哪些指标需要关注？

点击流分析的意义可分为两大方面：

1、技术上

- 可以合理修改网站结构及适度分配资源，构建后台服务器群组，比如
- 1、辅助改进网络的拓扑设计，提高性能
 - 2、在有高度相关性的节点之间安排快速有效的访问路径
 - 3、帮助企业更好地设计网站主页和安排网页内容

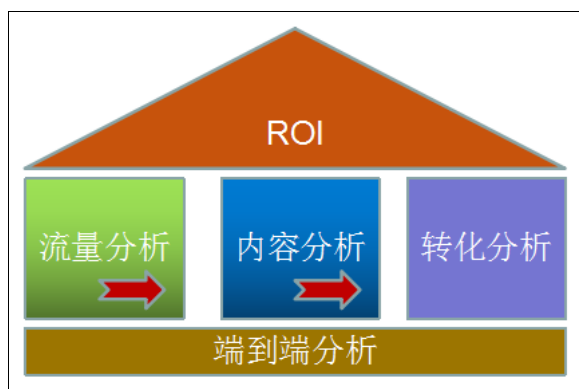
2、业务上

- 1、帮助企业改善市场营销决策，如把广告放在适当的 Web 页面上。
- 2、优化页面及业务流程设计，提高流量转化率。
- 3、帮助企业更好地根据客户的兴趣来安排内容。
- 4、帮助企业对客户群进行细分，针对不同客户制定个性化的促销策略等。

终极目标是：**改善网站(电商、社交、电影、小说)的运营，获取更高投资回报率（ROI）**

2.5、如何进行网站流量分析

流量分析整体来说是一个内涵非常丰富的体系，其整体过程是一个金字塔结构：



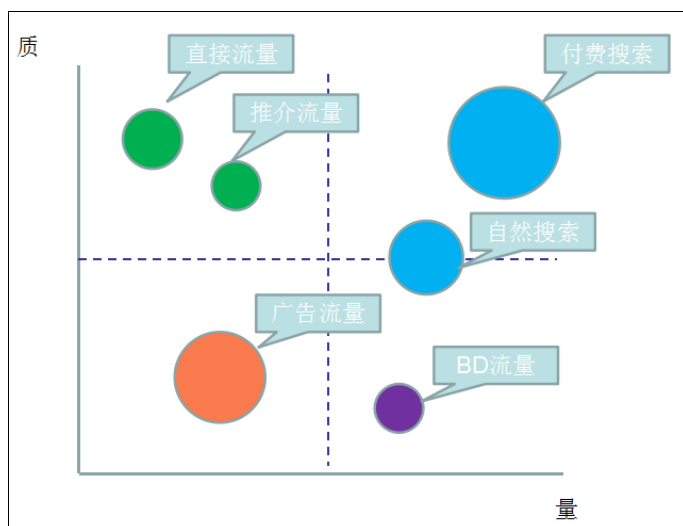
流量分析：分析量

内容分析：改善质

转化分析：提高投资回报率

2.5.1、网站流量质量分析

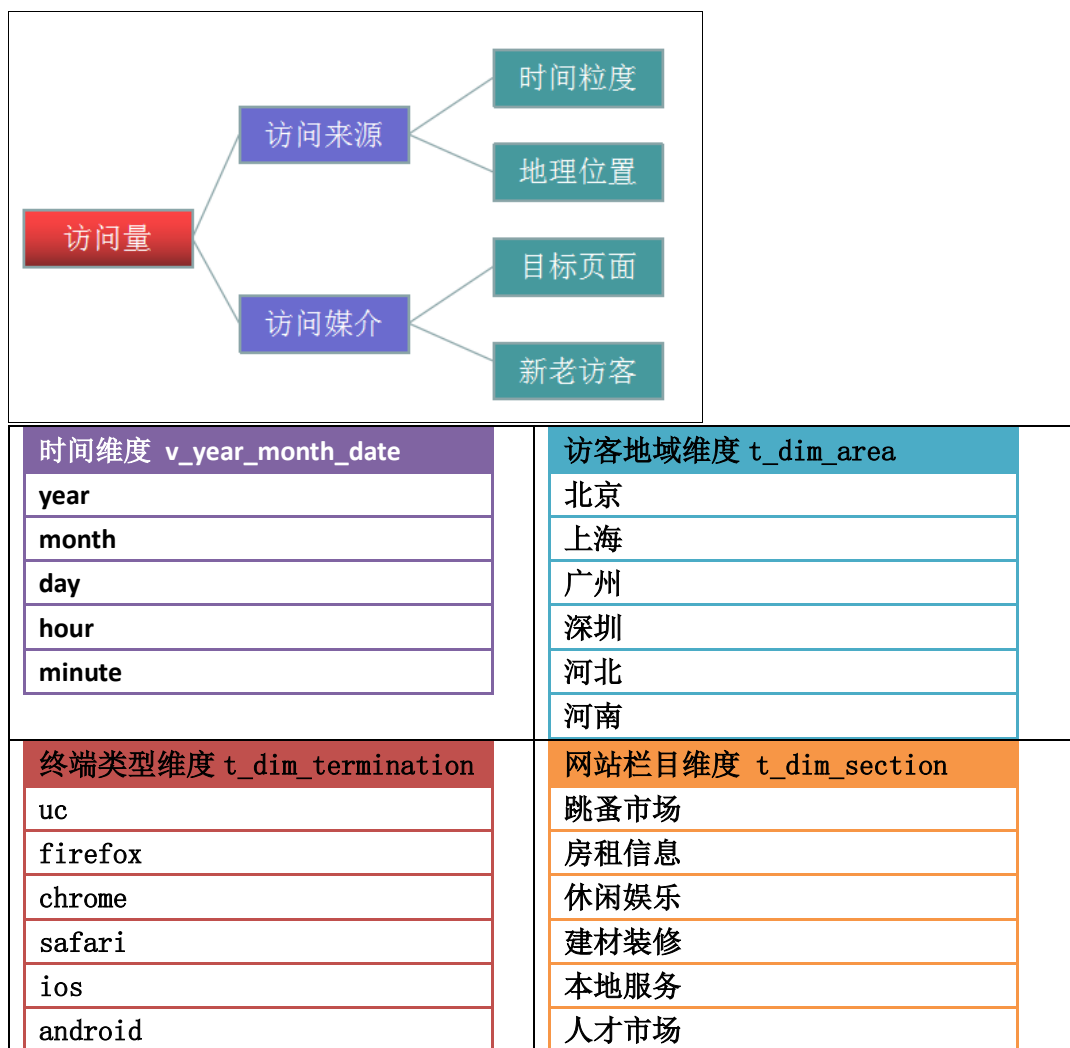
流量对于每个网站来说都是很重要，但流量并不是越多越好，应该更加看重流量的质量，换句话说就是流量可以为我们带来多少收入。



BD 流量：商务拓展流量

2.5.2、网站流量多维度细分

细分是指通过不同维度对指标进行分割，查看同一个指标在不同维度下的表现，进而找出有问题的那部分指标，对这部分指标进行优化。



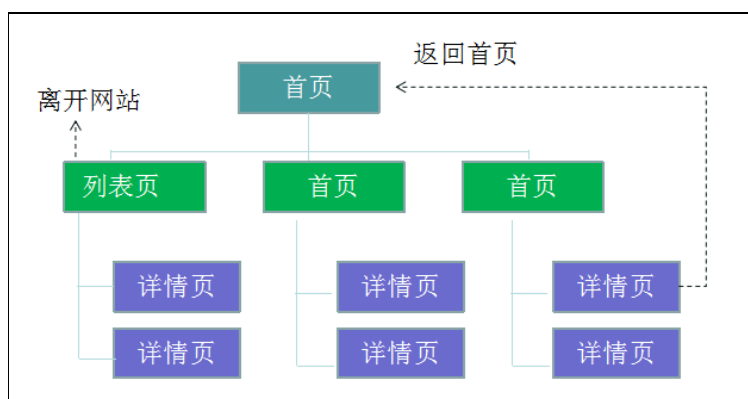
2.5.3、网站内容及导航分析

对于所有网站来说，页面都可以被划分为三个类别：

导航页 -- 首页和列表页都是典型的导航页；

功能页 -- 站内搜索页面、注册表单页面和购物车页面都是典型的功能页；

内容页 -- 比如从内容导航分析中，以下两类行为就是网站运营者不希望看到的行为：



第一个问题：访问者从导航页进入，在还没有看到内容页面之前就从导航页离开网站，需要

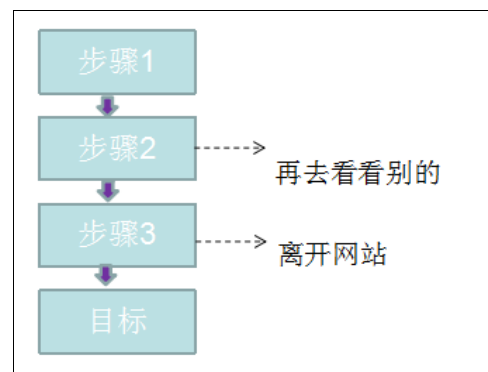
分析导航页造成访问者中途离开的原因。

第二个问题：访问者从导航页进入内容页后，又返回到导航页，说明需要分析内容页的最初设计，并考虑中内容页提供交叉的信息推荐

2.5.4、网站转化及漏斗分析

所谓转化，即网站业务流程中的一个封闭渠道，引导用户按照流程最终实现业务目标（比如商品成交）；而漏斗模型则是指进入渠道的用户在各环节递进过程中逐渐流失的形象描述；对于转化渠道，主要进行两部分的分析：访问者的流失和迷失

1、阻力和流失

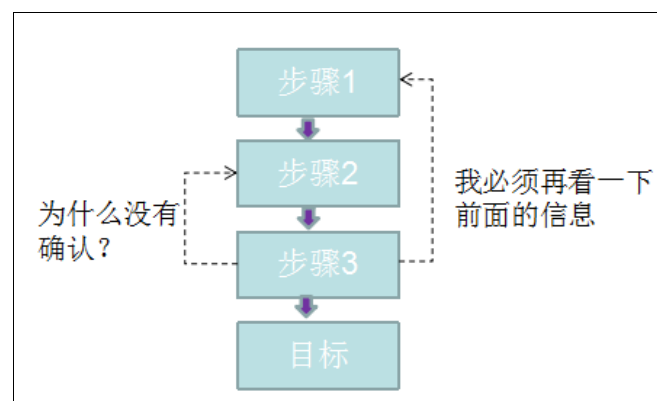


造成流失的原因很多，如：

不恰当的商品或活动推荐

对支付环节中专业名词的解释、帮助信息等内容不当

2、迷失



造成迷失的主要原因是转化流量设计不合理，访问者在特定阶段得不到需要的信息，并且不能根据现有的信息作出决策

总之，网站数据分析是一门内容非常丰富的学科，本课程中主要关注网站流量分析过程中的技术运用，更多关于网站数据分析的业务知识可学习《网站分析实战》

3、流量分析常见指标

3.1、基础分析

比如一些非常核心的指标：PV，IP，UV，独立访客

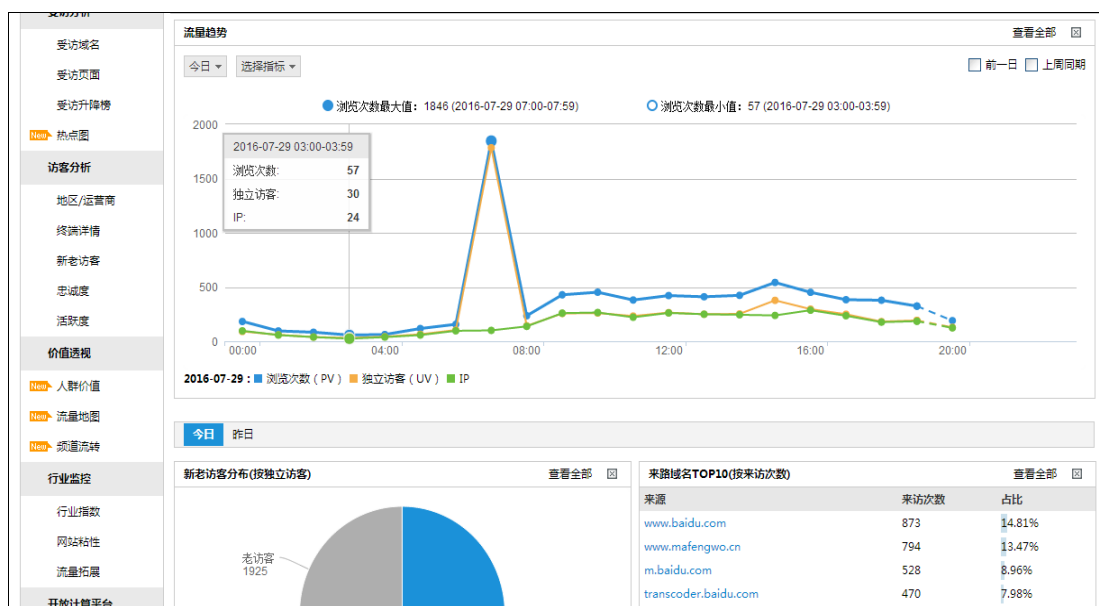
1、**趋势分析**：根据选定的时段，提供网站流量数据，通过流量趋势变化形态，为您分析网站访客的访问规律、网站发展状况提供参考。

2、**对比分析**：根据选定的两个对比时段，提供网站流量在时间上的纵向对比报表，帮您发现网站发展状况、发展规律、流量变化率等。

3、**当前在线**：提供当前时刻站点上的访客量，以及最近 15 分钟流量、来源、受访、访客变化情况，方便用户及时了解当前网站流量状况。

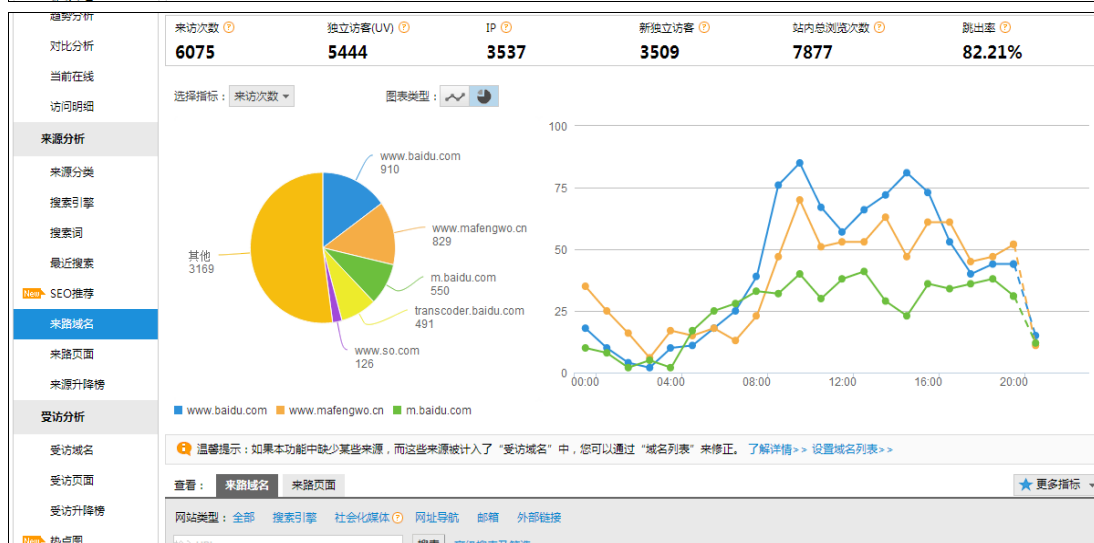
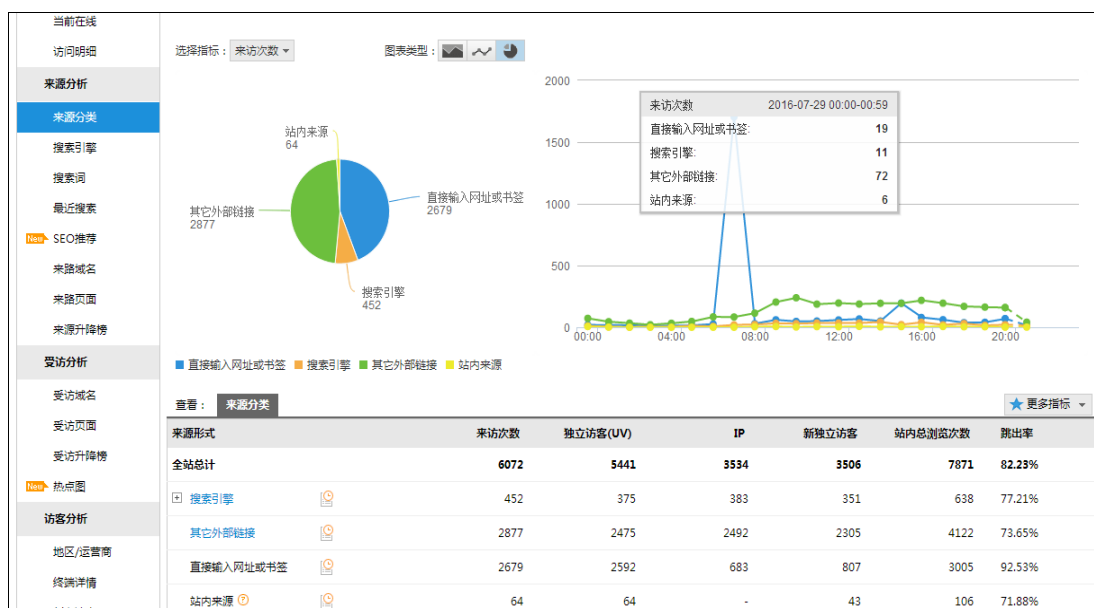
4、**访问明细**：提供最近 7 日的访客访问记录，可按每个 PV 或每次访问行为（访客的每次会话）显示，并可按照来源、搜索词等条件进行筛选。通过访问明细，用户可以详细了解网站流量的累计过程，从而为用户快速找出流量变动原因提供最原始、最准确的依据。

友盟+ 首页 产品 指数 报告 开发者中心 论坛 活动 注册 登录																																																											
CNZZ 数据专家																																																											
全站分析 无线流量分析 增值服务 信+金融																																																											
网站概况 (2016-07-29)																																																											
统计开通日期: 2015-08-04 CNZZ网站排名: 121200 查看Alexa排名																																																											
由于设置了过滤规则, 今日已过掉流量: 5PV 如果您对过滤数据有疑问, 可进行检查																																																											
<table> <tr> <th></th><th>浏览次数(PV)</th><th>独立访客(UV)</th><th>IP</th><th>新独立访客</th><th>访问次数</th></tr> <tr> <td>今日</td><td>7617</td><td>5290</td><td>3386</td><td>3365</td><td>5893</td></tr> <tr> <td>昨日</td><td>6229</td><td>3847</td><td>3697</td><td>3568</td><td>4549</td></tr> <tr> <td>今日预计</td><td>9690</td><td>6583</td><td>4177</td><td>4178</td><td>7401</td></tr> <tr> <td>昨日此时</td><td>5006</td><td>3158</td><td>3023</td><td>2903</td><td>3707</td></tr> <tr> <td>近90日平均</td><td>7283</td><td>4910</td><td>3732</td><td>3780</td><td>5737</td></tr> <tr> <td>历史最高</td><td>23280</td><td>21352</td><td>7434</td><td>21148</td><td>22130</td></tr> <tr> <td></td><td>(2016-02-18)</td><td>(2016-02-18)</td><td>(2015-08-19)</td><td>(2016-02-18)</td><td>(2016-02-18)</td></tr> <tr> <td>历史累计</td><td>2329562</td><td>1563408</td><td>1170289</td><td>-</td><td>-</td></tr> </table>							浏览次数(PV)	独立访客(UV)	IP	新独立访客	访问次数	今日	7617	5290	3386	3365	5893	昨日	6229	3847	3697	3568	4549	今日预计	9690	6583	4177	4178	7401	昨日此时	5006	3158	3023	2903	3707	近90日平均	7283	4910	3732	3780	5737	历史最高	23280	21352	7434	21148	22130		(2016-02-18)	(2016-02-18)	(2015-08-19)	(2016-02-18)	(2016-02-18)	历史累计	2329562	1563408	1170289	-	-
	浏览次数(PV)	独立访客(UV)	IP	新独立访客	访问次数																																																						
今日	7617	5290	3386	3365	5893																																																						
昨日	6229	3847	3697	3568	4549																																																						
今日预计	9690	6583	4177	4178	7401																																																						
昨日此时	5006	3158	3023	2903	3707																																																						
近90日平均	7283	4910	3732	3780	5737																																																						
历史最高	23280	21352	7434	21148	22130																																																						
	(2016-02-18)	(2016-02-18)	(2015-08-19)	(2016-02-18)	(2016-02-18)																																																						
历史累计	2329562	1563408	1170289	-	-																																																						
<table> <tr> <th></th><th>当日回头访客(占比)</th><th>访客平均访问频度</th><th>平均访问时长</th><th>平均访问深度</th><th>人均浏览页数</th></tr> <tr> <td>今日</td><td>405(7.66%)</td><td>1.11</td><td>59.49秒</td><td>1.29</td><td>1.44</td></tr> <tr> <td>昨日</td><td>489(12.71%)</td><td>1.18</td><td>1分13秒</td><td>1.37</td><td>1.62</td></tr> </table>							当日回头访客(占比)	访客平均访问频度	平均访问时长	平均访问深度	人均浏览页数	今日	405(7.66%)	1.11	59.49秒	1.29	1.44	昨日	489(12.71%)	1.18	1分13秒	1.37	1.62																																				
	当日回头访客(占比)	访客平均访问频度	平均访问时长	平均访问深度	人均浏览页数																																																						
今日	405(7.66%)	1.11	59.49秒	1.29	1.44																																																						
昨日	489(12.71%)	1.18	1分13秒	1.37	1.62																																																						
统计数据部件																																																											



3.2、来源分析

- 来源分类:** 提供不同来源形式（直接输入、搜索引擎、其他外部链接、站内来源）、不同来源项引入流量的比例情况。通过精确的量化数据，帮助用户分析什么类型的来路产生的流量多、效果好，进而合理优化推广方案。
- 搜索引擎:** 提供各搜索引擎以及搜索引擎子产品引入流量的比例情况。从搜索引擎引入流量的角度，帮助用户了解网站的 SEO、SEM 效果，从而为制定下一步 SEO、SEM 计划提供依据。
- 搜索词:** 提供访客通过搜索引擎进入网站所使用的搜索词，以及各搜索词引入流量的特征和分布。帮助用户了解各搜索词引入流量的质量，进而了解访客的兴趣关注点、网站与访客兴趣点的匹配度，为优化 SEO 方案及 SEM 提词方案提供详细依据。
- 最近 7 日的访客搜索记录:** 可按每个 PV 或每次访问行为（访客的每次会话）显示，并可按照访客类型、地区等条件进行筛选。为您搜索引擎优化提供最详细的原始数据。
- 来路域名:** 提供具体来路域名引入流量的分布情况，并可按“社会化媒体”、“搜索引擎”、“邮箱”等网站类型对来源域名进行分类。帮助用户了解哪类推广渠道产生的流量多、效果好，进而合理优化网站推广方案。
- 来路页面:** 提供具体来路页面引入流量的分布情况。尤其对于通过流量置换、包广告位等方式从其他网站引入流量的用户，该功能可以方便、清晰地展现广告引入的流量及效果，为优化推广方案提供依据。
- 来源升降榜:** 提供开通统计后任意两日的 TOP100 搜索词、来路域名引入流量的对比情况，并按照变化的剧烈程度提供排行榜。用户可通过此功能快速找到哪些来路对网站流量的影响比较大，从而及时排查相应来路问题。



访客分析		来路域名	来访次数	独立访客(UV)	IP	新独立访客	站内总浏览次数	跳出率
地区/运营商		全站总计	6075	5444	3537	3509	7877	82.21%
终端详情		趋势比较						
新老访客		www.baidu.com	910	740	756	690	1438	67.69%
忠诚度		www.mafengwo.cn	829	813	812	757	1188	68.76%
活跃度		m.baidu.com	550	453	459	424	712	82.91%
价值透视		http://m.baidu.com/from...	3	1	1	1	7	33.33%
人群价值		http://m.baidu.com/from...	3	1	1	1	3	100%
流量地图		http://m.baidu.com/from...	3	1	1	1	4	66.67%
频道流转		http://m.baidu.com/from...	3	1	1	1	3	100%
行业监控		http://m.baidu.com/from...	3	1	1	1	3	100%
行业指数		http://m.baidu.com/from...	3	1	2	1	3	100%
网站粘性		http://m.baidu.com/from...	3	1	1	1	4	66.67%
流量拓展		http://m.baidu.com/from...	2	1	1	1	3	50%
开放计算平台		http://m.baidu.com/from...	2	1	1	1	2	100%
事件		http://m.baidu.com/from...	2	1	1	1	2	100%
事件类别		https://m.baidu.com/fro...	2	1	1	1	3	50%
发生页面		transcoder.baidu.com	491	382	393	355	652	79.63%

3.3、受访分析

- 1、受访域名：**提供访客对网站中各个域名的访问情况。一般情况下，网站不同域名提供的产品、内容各有差异，通过此功能用户可以了解不同内容的受欢迎程度以及网站运营成效。
- 2、受访页面：**提供访客对网站中各个页面的访问情况。站内入口页面为访客进入网站时浏览的第一个页面，如果入口页面的跳出率较高则需要关注并优化；站内出口页面为访客访问网站的最后一个页面，对于离开率较高的页面需要关注并优化。
- 3、受访升降榜：**提供开通统计后任意两日的 TOP100 受访页面的浏览情况对比，并按照变化的剧烈程度提供排行榜。可通过此功能验证经过改版的页面是否有流量提升或哪些页面有巨大流量波动，从而及时排查相应问题。
- 4、热点图：**记录访客在页面上的鼠标点击行为，通过颜色区分不同区域的点击热度；支持将一组页面设置为"关注范围"，并可按来路细分点击热度。通过访客在页面上的点击量统计，可以了解页面设计是否合理、广告位的安排能否获取更多佣金等。
- 5、用户视点：**提供受访页面对页面上链接的其他站内页面的输出流量，并通过输出流量的高低绘制热度图，与热点图不同的是，所有记录都是实际打开了下一页面产生了浏览次数（PV）的数据，而不仅仅是拥有鼠标点击行为。
- 6、访问轨迹：**提供观察焦点页面的上下游页面，了解访客从哪些途径进入页面，又流向了哪里。通过上游页面列表比较出不同流量引入渠道的效果；通过下游页面列表了解用户的浏览习惯，哪些页面元素、内容更吸引访客点击。

对比分析	查看: 来访页面	2016-07-28	2016-07-27	+升 -降 平 全部
当前在线				
访问明细				
来源分析	全站总计	6229	7928	-1699(-21.43%)
来源分类	http://huoche.mafengwo.cn/	980	1084	-104(-9.59%)
搜索引擎	http://huoche.mafengwo.cn/skb/c-2189.html	5	25	-20(-80%)
搜索词	http://huoche.mafengwo.cn/skb/z-436.html	99	81	+18(22.22%)
最近搜索	http://huoche.mafengwo.cn/skb/	66	83	-17(-20.48%)
SEO推荐	http://huoche.mafengwo.cn/skb/z-45.html	3	18	-15(-83.33%)
来源域名	http://huoche.mafengwo.cn/skb/c-3608.html	18	4	+14(350%)
来源页面	http://huoche.mafengwo.cn/skb/zz-113-741.html	0	14	-14(-100%)
来源升降梯	http://huoche.mafengwo.cn/skb/z-525.html	37	24	+13(54.17%)
受访问分析	http://huoche.mafengwo.cn/skb/z-74.html	11	24	-13(-54.17%)
受访问域名	http://huoche.mafengwo.cn/skb/z-527.html	16	5	+11(220%)
受访问页面	http://huoche.mafengwo.cn/skb/zz-1-6.html	11	0	+11(-)
受访问升降梯	http://huoche.mafengwo.cn/ysq/	21	32	-11(-34.38%)
热点图	http://huoche.mafengwo.cn/skb/zz-86-373.html	13	3	+10(333.33%)
访客分析	http://huoche.mafengwo.cn/skb/c-563.html	4	14	-10(-71.43%)
地区/运营商				

对比分析
当前在线
访问明细
来源分析
来源分类
搜索引擎
搜索词
最近搜索
SEO推荐
来源域名
来源页面
来源升降梯
受访问分析
受访问域名
受访问页面
受访问升降梯
热点图
访客分析
地区/运营商

热点图名称

统计范围

显示页面

统计模式

查看热点图

编辑

推荐使用Chrome浏览器查看热点图，支持在Firefox、IE8及更高版本下查看。
如果是购物车、订单详情页这种需要登录才可查看的网页，则可在您的网站登录，再查看对应热点图。



热点图功能记录页面访客的鼠标点击行为，通过颜色区分不同区域点击热度。支持“关注范围”的设置，按来源细分点击热度等功能。

页面热点，访客兴趣尽在掌握。
页面优化、广告位规划轻松完成。

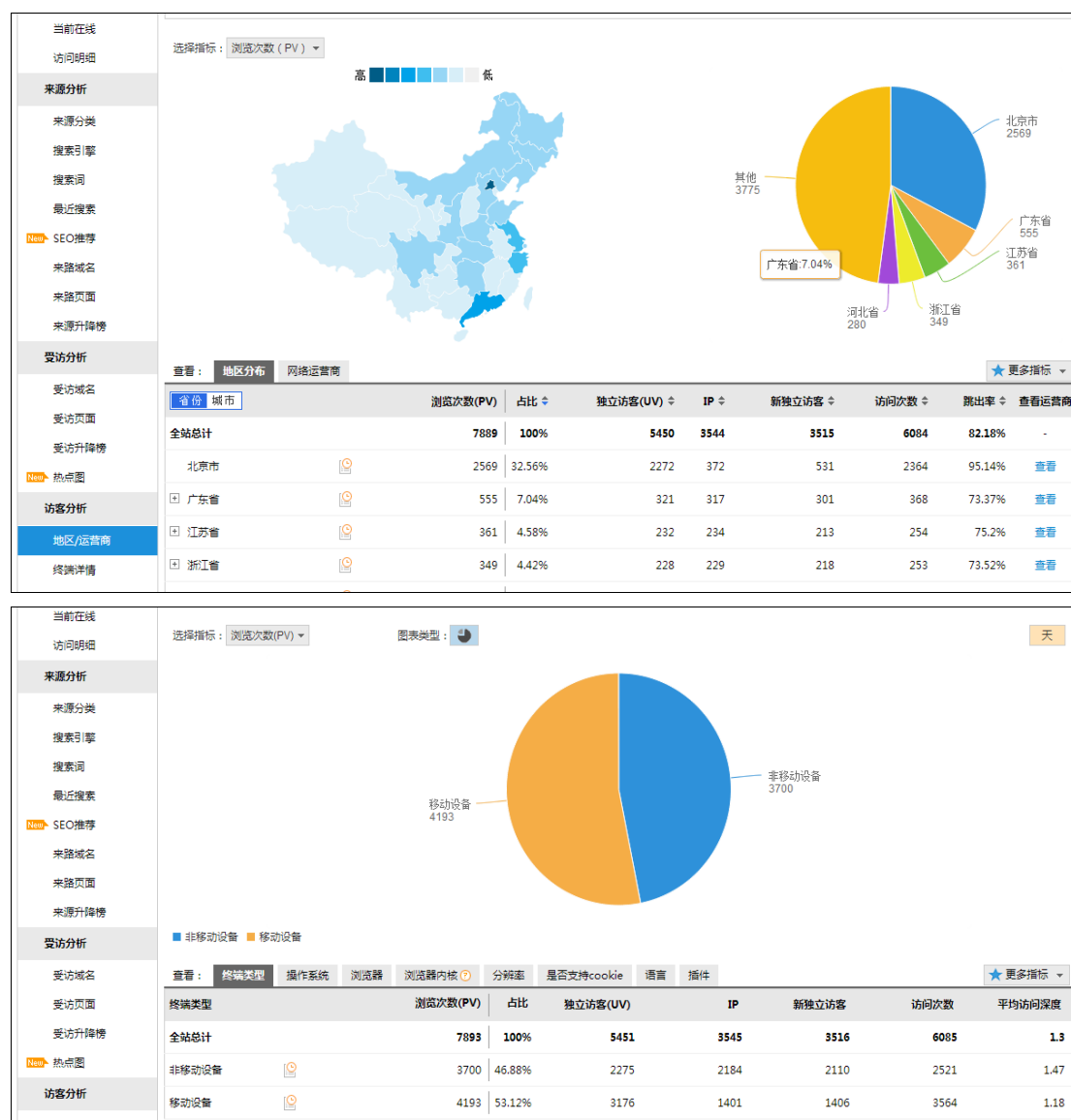
+ 立即添加热点图

3.4、访客分析

- 地区运营商**：提供各地区访客、各网络运营商访客的访问情况分布。地方网站、下载站等与地域性、网络链路等结合较为紧密的网站，可以参考此功能数据，合理优化推广运营方案。
- 终端详情**：提供网站访客所使用的浏览终端的配置情况。参考此数据进行网页设计、开发，可更好地提高网站兼容性，以达到良好的用户交互体验。
- 新老访客**：当日访客中，历史上第一次访问该网站的访客记为当日新访客；历史上已经访问过该网站的访客记为老访客。新访客与老访客进入网站的途径和浏览行为往往存在差异。该功能可以辅助分析不同访客的行为习惯，针对不同访客优化网站，例如为制作新手导航提供数据支持等。
- 忠诚度**：从访客一天内回访网站的次数（日访问频度）与访客上次访问网站的时间两个角度，分析访客对网站的访问粘性、忠诚度、吸引程度。由于提升网站内容的更新频率、增强用户体验与用户价值可以有更高的忠诚度，因此该功能在网站内容更新及用户体验方面提

供了重要参考。

5、**活跃度**：从访客单次访问浏览网站的时间与网页数两个角度，分析访客在网站上的活跃程度。由于提升网站内容的质量与数量可以获得更高的活跃度，因此该功能是网站内容分析的关键指标之一。



来源页面	新访客	5739	浏览次数	2157
来源升降榜				
受访问分析		3517	独立访客	1935
受访问域名		4090	访问次数	1997
受访问页面		1.63	人均浏览页数	1.11
受访问升降榜				
热点图		1.4	平均访问深度	1.08
访客分析		58.49秒	平均访问时长	1分7秒
地区/运营商		75.33%	跳出率	96.19%
终端详情				
新老访客				
忠诚度				
活跃度				
价值透视				
人群价值				
流量地图				
频道流转				

新访客来源TOP5(按独立访客)			老访客来源TOP5(按独立访客)		
来源	独立访客	站内总浏览次数	来源	独立访客	站内总浏览次数
www.mafengwo.cn	757	1108	www.mafengwo.cn	56	80
www.baidu.com	691	1339	www.baidu.com	50	107
m.baidu.com	425	670	m.baidu.com	29	44
transcoder.baidu.com	356	574	transcoder.baidu.com	27	80
www.so.com	102	178	wap.sogou.com	4	12

网站概况	访客分析-忠诚度 (2016-07-29至2016-07-29)	下载报表
流量分析	<div> < > 今日 昨日 前日 2016-07-29 </div>	
来源分析	独立访客(UV)	浏览次数(PV)
受访问分析	5603	8125
访客分析	查看: 日访问频度 上次访问时间	
地区/运营商	日访问频度	独立访客(UV) 占比 浏览次数(PV) 占比
终端详情	只访问1次	5160 92.09% 6393 78.68%
新老访客	2次	320 5.71% 979 12.05%
忠诚度	3次	78 1.39% 360 4.43%
活跃度	4次	24 0.43% 158 1.94%
价值透视	5次	10 0.18% 70 0.86%
人群价值	6次及以上	11 0.2% 165 2.03%
流量地图		
频道流转		
行业监控		

访客分析	查看: 访问时长分布 访问深度分布	
地区/运营商	访问时长	访问次数 占比 浏览次数(PV) 占比
终端详情	小于10秒	311 4.96% 639 7.86%
新老访客	11-30秒	364 5.8% 845 10.39%
忠诚度	31-60秒	167 2.66% 407 5%
活跃度	1-3分钟	175 2.79% 577 7.1%
价值透视	3-10分钟	62 0.99% 331 4.07%
人群价值	11-30分钟	11 0.18% 99 1.22%
流量地图	30分钟以上	1 0.02% 15 0.18%
频道流转	无时长记录	5037 80.3% 5041 61.99%
行业监控		

3.5、转化路径分析

转化定义

1、访客在您的网站完成了某项您期望的活动，记为一次转化，如注册或下载。

目标示例

- 1、获得用户目标：在线注册、创建账号等。
- 2、咨询目标：咨询、留言、电话等。
- 3、互动目标：视频播放、加入购物车、分享等。

4、收入目标：在线订单、付款等。

转化数据的应用

- 1、在报告的自定义指标中勾选转化指标，实时掌握网站的推广及运营情况。
- 2、结合“全部来源”、“转化路径”、“页面上下游”等报告分析访问漏斗，提高转化率。
- 3、对“转化目标”设置价值，预估转化收益，衡量 ROI。

路径分析：根据设置的特定路线，监测某一流程的完成转化情况，算出每步的转换率和流失率数据，如注册流程，购买流程等。

转化类型：

1、页面



2、事件

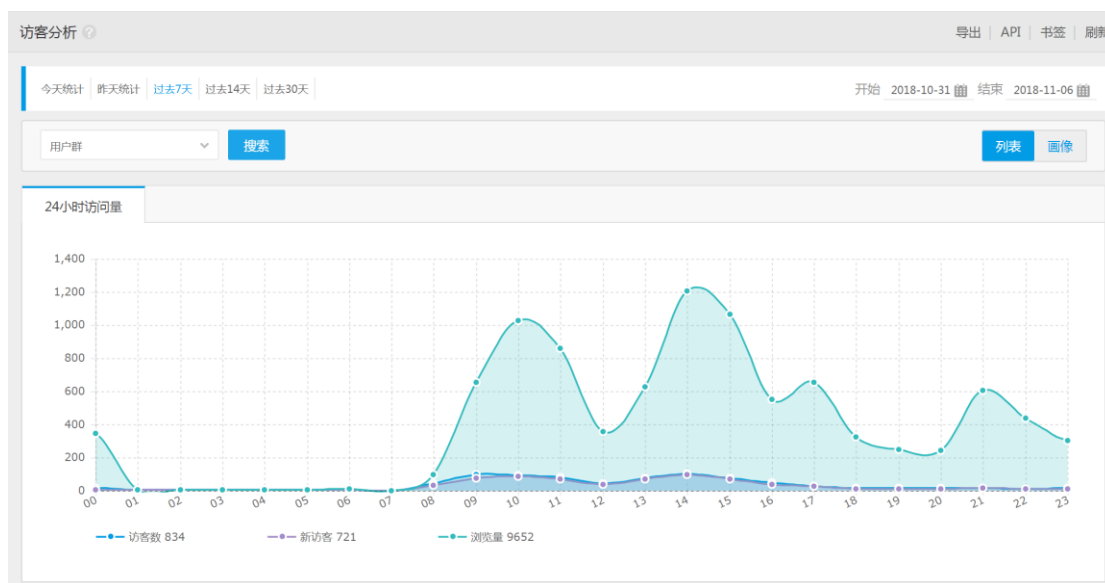


3.6、用户分析模块

主要分析**新增会员**（如果按天分析，就是昨天注册的会员）、**活跃会员**（只要今天登陆的就是活跃会员，或者按照一定的周期去计算，只要在这个周期内都是活跃会员）以及**总会员**相关信息

1、访客：

主要分析**新增用户**、**活跃用户**以及**总用户**的相关信息



- 2、会员：主要分析**新增会员**、**活跃会员**以及**总会员**的相关信息
- 3、会话：主要分析**会话个数**，**会话长度**和**平均会话长度**相关的信息
- 4、主要分析每天每小时的用户、会话个数以及会话长度的相关信息

3.7、浏览器分析模块

每个浏览器的平均会话访问时间

每个浏览器每天的会话个数

在用户基本信息分析的基础上添加一个浏览器这个维度信息

包括：

- 浏览器用户分析
- 浏览器会员分析
- 浏览器会话分析
- 浏览器 PV 分析

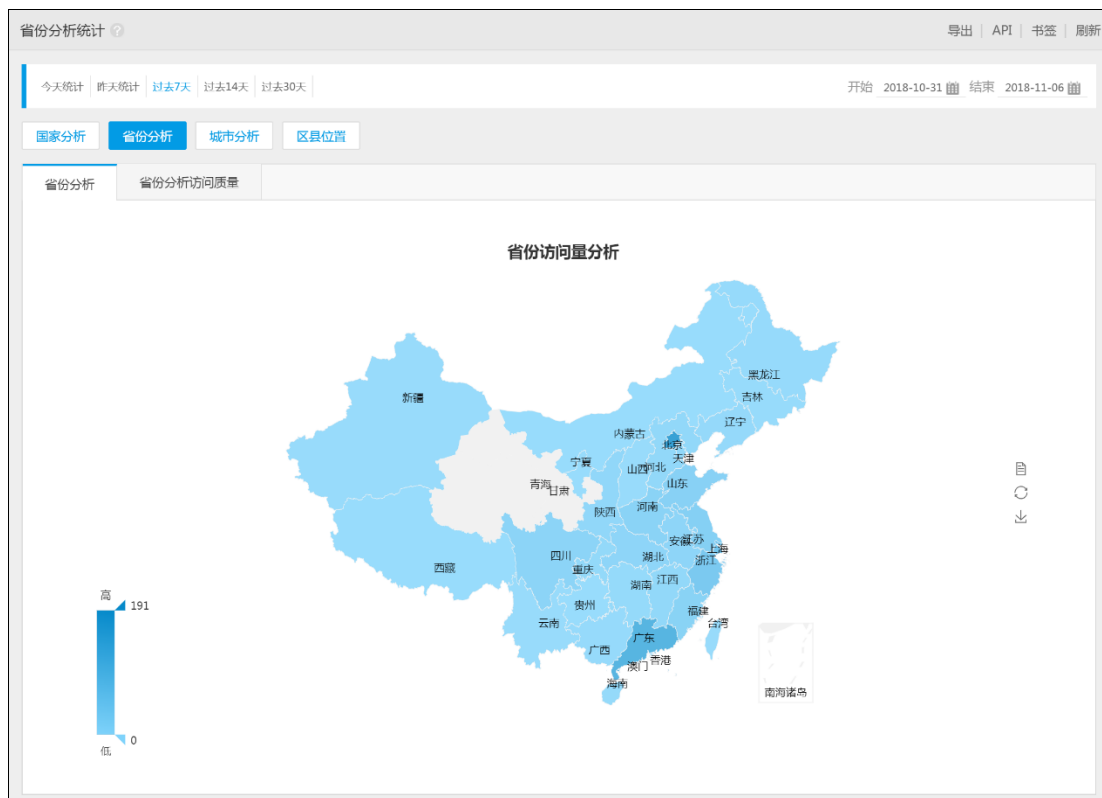
3.8、地域分析模块

主要分析各个不同省份的用户和会员情况

包括：

活跃访客地域分析

不同地域的跳出率分析



3.9、用户浏览深度分析模块

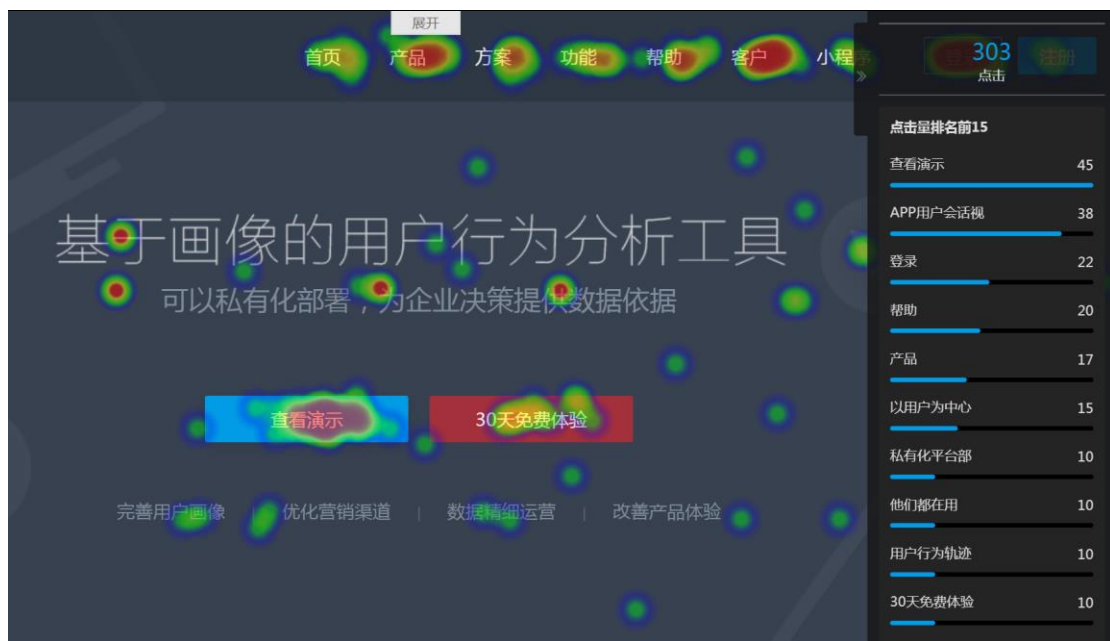
记录的是每天每个用户或者会员访问了多少个页面。

3.10、订单分析模块

主要分析订单的相关情况：订单数，成功支付订单，退款订单，支付失败订单占比

3.11、热力图分析

主要用来分析用户在访问一个重要页面的时候鼠标主要停留在什么位置，以此分析出用户最感兴趣的内容。



3.12、投入产出分析（ROI）

主要分析投入产出比

4、项目技术架构

4.1、数据处理流程

该项目是一个纯粹的数据分析项目，其整体流程基本上就是依据数据的处理流程进行，依此有以下几个大的步骤：

1、数据采集

首先，通过页面嵌入 JS 代码的方式获取用户访问行为，并发送到 web 服务的后台记录日志。然后，将各服务器上生成的点击流日志通过实时或批量的方式汇聚到 HDFS 文件系统中。当然，一个综合分析系统，数据源可能不仅包含点击流数据，还有数据库中的业务数据（如用户信息、商品信息、订单信息等）及对分析有益的外部数据。

2、数据预处理

通过 mapreduce 程序对采集到的点击流数据进行预处理，比如清洗，格式整理，滤除脏数据等。

3、数据入库

将预处理之后的数据导入到 HIVE 仓库中相应的库和表中。

4、数据分析

项目的核心内容，即根据需求开发 ETL 分析语句，得出各种统计结果

5、数据展现

将分析所得数据进行可视化

4.2、技术选型

4.2.1、数据收集层

数据收集层涉及到写数据埋点，数据埋点可分为两类：

A：前台数据埋点：

使用 JavaScript 去写

问题：为什么不用 JQuery 写

B：后台数据埋点：

使用 Java 去写

Flume：收集日志（如果实时的还需要 Kafka）

MapReduce：对数据进行预处理

4.2.2、数据分析层

MapReduce

Hive

Kylin/pig/impala/drill

SparkSQL

SparkCore

Oozie/azkaban/crontab

Hive + HBase(SQL)

HBase + Phoenix

4.2.3、数据展示层

Sqoop 导入到 MySQL 或是 Oracle/PostGre(关系型数据库)/HBase

结合三大框架：

Spring + Struts2 + MyBatis ==> SSM ss2m

Spring + Struts2 + hibernate ==> SSH ss2h

Spring + SpringMVC + MyBatis

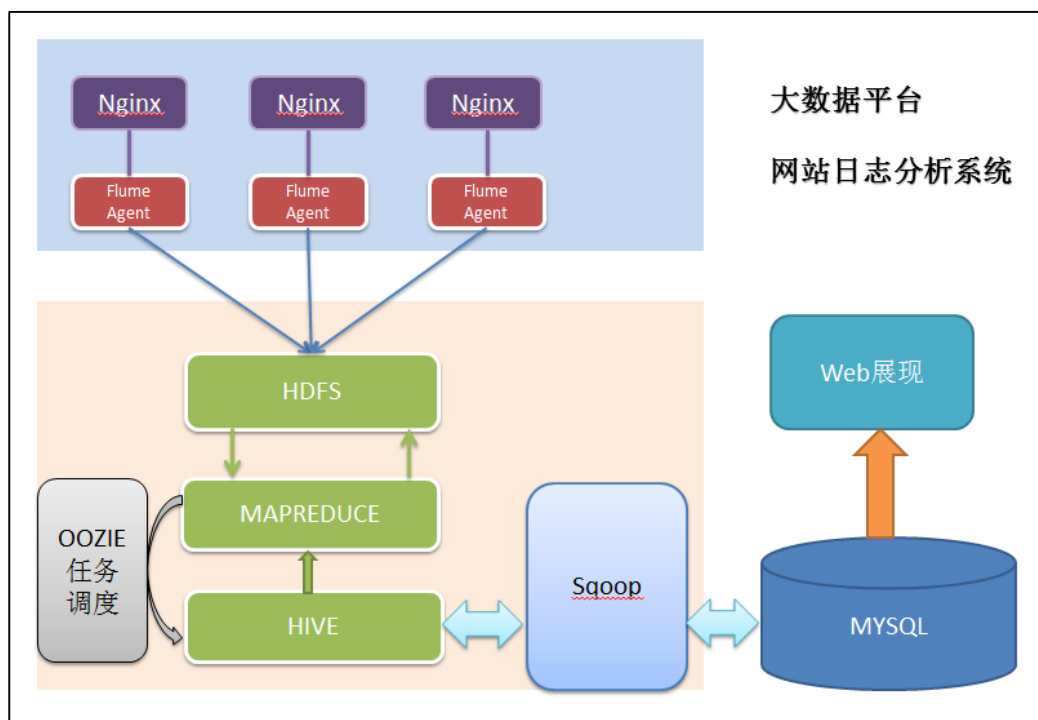
SpringBoot + Spring + MyBatis 等

Echarts/Highcharts

Tableau、Bootstrap、EasyUI、D3.js

4.3、项目结构

由于本项目是一个纯粹数据分析项目，其整体结构亦跟分析流程匹配，并没有特别复杂的结构，如下图：



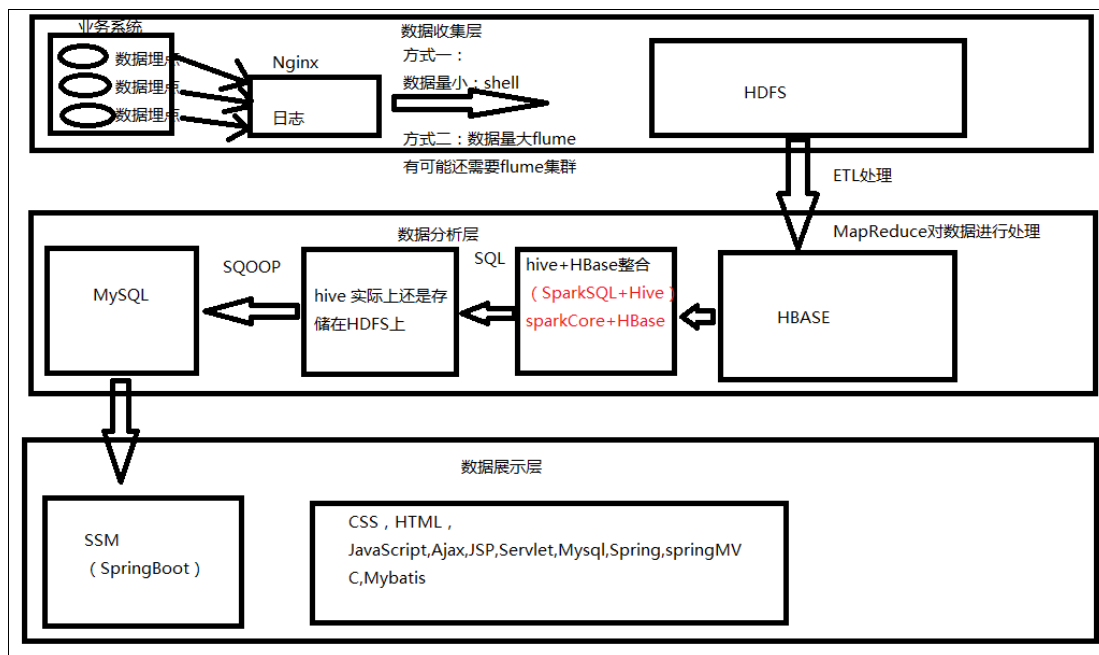
其中，需要强调的是：

系统的数据分析不是一次性的，而是按照一定的时间频率反复计算，因而整个处理链条中的各个环节需要按照一定的先后依赖关系紧密衔接，即涉及到大量任务单元的管理调度，所以，项目中需要添加一个任务调度模块

4.4、总体架构分析

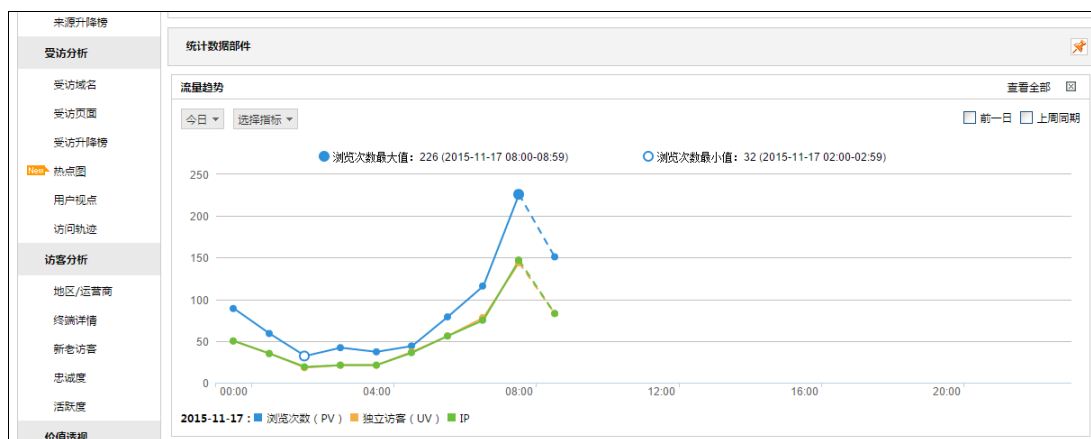
基本上，这一套架构几乎是离线架构的万能架构。总体上分为以下三个部分：

- 1、数据分析层
- 2、数据处理层
- 3、数据展示层



4.5、数据展现

数据展现的目的是将分析所得的数据进行可视化，以便运营决策人员能更方便地获取数据，更快更简单地理解数据



5、集群开发环境准备

5.1、集群规划

	Hadoop02	Hadoop03	Hadoop04	Hadoop05
ZooKeeper	√	√	√	

NameNode	√	√		
DataNode	√	√	√	√
JournalNode	√	√	√	
ZKFC	√	√		
ResourceManager			√	√
NodeManager	√	√	√	√
Nginx				√
Flume				√
MySQL	√			
HMaster	√		√	
HRegionServer	√	√	√	√

5.2、环境搭建

- 1、安装 JDK
- 2、安装 Nginx
- 3、安装 MySQL
- 4、安装 ZooKeeper
- 5、安装 Hadoop
- 6、安装 HBase
- 7、安装 Hive
- 8、安装 Flume
- 9、安装 Azkaban
- 10、安装 Sqoop