业务背景

网站流量统计是改进网站服务的重要手段之一，通过获取用户在网站的行为数据，进行分析,得到有价值的信息。可以基于这些数据对网站进行改进。

按照在线情况分析：

来访时间、访客地域、来路页面、当前停留页面等，这些功能对企业实时掌握自身网站流量有很大的帮助。

按照时段分析：

按照任意的时间段 和任意的时间粒度来进行分析，比如小时段分布，日访问量分布，对于企业了解用户浏览网页的的时间段有一个很好的分析。

按来源分析

来源分析提供来路域名带来的来访次数、IP、独立访客、新访客、新访客浏览次数、站内总浏览次数等数据。这个数据可以直接让企业了解推广成效的来路，从而分析出那些网站投放的广告效果更明显。

业务需求

PV

PV(page view)，访问量 也叫 点击量，即一天之内整个网站中的页面被访问的次数。对同一个页面的重复访问记为不同的PV。

UV

UV(unique visitor),独立访客数，即一天之内访问网站的人数。同一个人在一天之内访问网站多次，也只能计算一个UV。

VV

VV (Visit View)，会话总数，即一天之内会话的总的数量。所谓的一次会话，指的是为了实现某些功能，浏览器开发网站网站，从访问第一个页面开始，会话开始，直到访问最后一个页面结束，关闭所有页面，会话结束。会话可以认为在访问第一个页面时开始，访问所有页面完成并关闭，或，超过指定时长没有后续访问 都认为会话结束。

BR

BR(Bounce Rate)跳出率，一天之内跳出的会话总数占所有会话总数的比率。所谓跳出只的是一个会话中只访问过一个页面会话就结束了，这就称之为该会话跳出了。跳出的会话占全部会话的比率，称之为跳出率。这个指标在评价行推广活动的效果时非常的有用。

NewIP

NewIp,新增ip总数,一天之内访问网站的所有IP去重后，检查有多少是在历史数据中从未出现过的，这些IP计数，就是新增的IP总数，这个指标可以一定程度上体现网站新用户增长的情况。

NewCust

NewCust，新增独立访客数，一天之内访问网站的人中，有多少人是在历史记录中从来没有出现过的。这个指标可以从另一个角度体现网站用户增长的情况。

AvgTime

AvgTime,平均访问时长，所谓一个会话的访问时长，是指，一个会话结束的时间减取一个会话开始时间得到的会话经历的时长。将一天之内所有会话的访问时长求平均值 就是 平均访问时长。这个指标可以体现出网站对用户的粘性大小。

AvgDeep

AvgDeep，平均访问深度，所谓一个会话的访问深度，是指，一个会话中访问的所有资源地址去重后计数得到的指标。将一天之内所有会话的访问深度求平均值 就是 平均访问深度。这个指标可以体现出网站对用户的粘性大小。

技术架构

计算机生成了可选文字:
日 吉 收 隼 
网 灾 埋 貞 
http //xxxx 
js 脚 本 收 隼 數 据 
Nglnx 
A 」 A × 提 交 數 钅 刂 服 务 器 
Nglnx 
应 用 服 务 器 
应 用 服 务 器 
应 用 服 务 器 
日 志 服 务 器 
应 用 服 务 器 
应 用 服 务 器 
Flume 
Agent 
Flume 
Agent 
Flume 
Agent 
网 站 流 量 分 折 顼 目 
数 据 存 储 
Kafka 
肖 寻 队 列 
HDFS 
分 布 式 存 储 
數 据 青 先 处 邏 
HBase 
HDFS 
HIVE 
数 据 仓 库 
Redis 
1-41Ve 
數 据 导 出 
M 
可 视 化 展 示 
http //xxxx 
折 线 图 
饼 图 
Flume 
中 心 服 务 器 
SparkStream1ng 
实 时 分 折 
分 布 式 计 算 