慕课网 网站分析

1. 日志：

日志的数据格式：

|  |
| --- |
| 194.237.142.21 - - [18/Sep/2013:06:49:18 +0000] "GET /wp-content/uploads/2013/07/rstudio-git3.png HTTP/1.1" 304 0 "-" "Mozilla/4.0 (compatible;)" |

日志的概念：

一般日志，指定就是网络设备、业务系统以及后台服务器程序等等，所产生的数据，它在运行过程中，一般产生一个叫做xxxx.log文件，该文件就是存储日志的文本文件；

一条完成的日志包含有：日期、时间、使用者所操作的相关信息数据；

在日志当中，最具有代表性的就是网站日志 ，

网站日志就是网站在运行过程中所产生的日志，包含系统日志和程序日志；

1. 系统日志：一般就是网站所基于的服务器所在自动生成的日志；

比如访问日志，系统的错误日志；

1. 程序日志：由程序开发的过程中，程序员自定义的日志；
2. 为什么要分析日志数据？

网站的日志本身并没有价值，只有当我们对其进行相关的处理、分析、挖掘之后产生会产生价值，才能够让原始的日志数据进行“说话”；

比如：在对于网站日志的分析中，我们可以了解：

1. 网站的运营的情况：

PV : 一天当中，网站的访问量（在一个时间段内，网站被访问了多少次）；

只要有一次跳转，就代表一次PV；

UV ：一天当中，网站内多少个用户访问过；（对网站来说，后台会ip来作为用户的唯一标识）

1. 网站的安全情况：

比如：恶意攻击，密码暴力破解；

恶意攻击：

爬虫（可以达到1秒内访问多次网站的效果，对于来说会监控到该访问的ip，）；

密码暴力破解：

木马（盗取密码，就前端页面把用户的账号和密码的数据，往后台发送时，木马病毒程序，会从中拦截发送的数据包，然后通过破解来获取到相关的密码信息）；

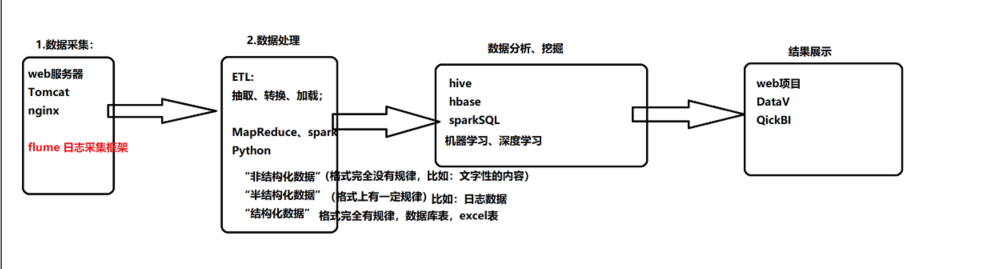
1. 网站的用户信息：

通过对用户日志数据的分析，可以获取到访问用户操作系统，浏览器等等一些数据，当获取到这些数据后，我们可以对网站的安全或者网站的广告投放进行相关的设计；

比如：京东，它会统计用户一般情况下，是如何来登录京东的。是从百度搜索、还是直接请求京东的网站域名，还是从其他的广告网站跳转过来的。

当京东获取到这些数据后，会进行分析，比如百度搜索来的用户比较多，那么京东就可以在百度搜索上进行更大的资金投入，让收益最大化；

1. 网站日志分析数据的流程：



1. 一个完整的网站访问日志包含：

|  |
| --- |
| 194.237.142.21 - - [18/Sep/2013:06:49:18 +0000] "GET /wp-content/uploads/2013/07/rstudio-git3.png HTTP/1.1" 304 0 "-" "Mozilla/4.0 (compatible;)" |

|  |
| --- |
| 1、用户的ip : 分城市和网络运营商的；ip和路由器有机密关系；  请问，我们当前教师访问一个网站时，所产生的ip有几个；  解释：只有一个，因为我们所有同学访问外网时，都是通过一个路由器来访问的，所以对网站来说，它能够获取到的用户ip仅仅是路由器的ip;   1. 访问时间：默认的时间显示格式是国际标准格式； 2. 请求链接：访问的网站的url , url中包含有你请求的该网站内的资源；   地 请问GET,和post有什么区别？  4、地方电 4、客户端信息：用户的操作系统，浏览器内核信息；  4 |

五、

分析1：网站的流量分析；

PV, UV IP ;

请问一个网站如果要统计流量时，是uv统计比较准确，还是IP统计比较准确？

解释：UV更准确一些；

分析2：网站来源分析：

用户从各个来源网站进入本网站的占有比例；

主要是的搜索引擎来源，来源中的各个搜索引擎的占有比例，比如：百度、谷歌，360；

分析3：网站的访问分析：

访问链接的TopN : 一个大型的网站肯定是由一系列的网页组成的，topN就是代表一个网站中，有哪些网页是最容易被用户的访问的。

恶意攻击的ip : 比如一个ip竟然可以在一秒内访问网站多次，而是不访问不同的网页的URL，那么这种ip肯定是爬虫程序；

分析4：网站访客分析：

主要分析访客的操作系统：window mac, 安卓（手机）；作用：可以了解到网站的访问渠道是pc端多，还是移动端多；

分析网络运营商信息，可以掌握访问当前网站的哪家运营商的网络比较多；

电信、联通（网最稳当）、移动（信号最稳当，基站比较多）；

1. 分析目的：
2. 网站程序可以更好的优化；
3. 网站服务器优化；
4. 搜索引擎可以更好得到推广；