

UNIT IV WEB ANALYTICS

Web Analytics - Present and Future, Data Collection - Importance and Options, Overview of Qualitative Analysis, Business Analysis, KPI and Planning, Critical Components of a Successful Web Analytics Strategy, Web Analytics Fundamentals, Concepts, Proposals & Reports, Web Data Analysis

WEB ANALYTICS - PRESENT AND FUTURE

Web analytics is the objective tracking, collection, measurement, reporting, and analysis of quantitative Internet data to optimize websites and web marketing initiatives.

History of Web Analytics

First log file analysis program was written by Dr. Stephen Turner in 1995. It was one of the First log file analysis programs that was widely available on the Web. It is still one of the most widely used web analytics applications and it comes installed on websites from most Internet Service Providers (ISPs). Analog, and tools like it, fueled the adoption of web analytics beyond the Information Technology (IT) team. The reports started to get prettier, and of course marketing folks could now finally understand what was happening.

Around 1995–96, the general users of the Internet started to get exposed to web statistics because of the proliferation of a delightful thing called a counter. **Page counters** were perhaps the first example of web viral marketing (credited to a company called Web-Counter). Counters were everywhere on the Web; they stood for both being cool and showing how popular you were.

Commercial web analytics started several years later. WebTrends is the standard log file parser which added tables and pretty graphs that finally dragged web analytics to the business teams.

By the year 2000, with the popularity of the Web growing exponentially, web analytics was established as a discipline. Companies such as Accrue, WebTrends, WebSideStory, and Coremetrics were all firmly established as key vendors, providing increasingly complex solutions that reported massive amounts of data.

Around the same time, web analytics vendors and customers were discovering that using web server logs as optimal sources of data presented certain challenges.

Challenges with using the logs:

1. Page Caching by ISP:

The challenge with caching is, if the ISP had a copy of the page, all subsequent pages would be served from the ISP, and the website log files would not have entries for those requested pages.

2. Search Robots:

With the increasing popularity of search engines, search bots would frequently crawl sites and leave non-web-user entries in web logs. These entries would be counted in the metrics.

3. Unique Visitors:

With an increasing number of users with dynamic IP addresses and the addresses coming via proxy servers, it became difficult to identify unique visitors. Vendors resorted to using the IP address plus the user agent ID (user operating system and browser), but that was not optimal. To overcome these challenges Javascript tags were used.

The four big vendors of web analytics are: Coremetrics, Omniture, WebTrends, and WebSideStory. There are also a whole host of mid-market vendors such as Unica, indexTools, and ClickTracks, and many basic solutions such as the open source products AWStats, Webalizer, and StatCounter.

Google had a major effect on the web analytics landscape in 2005 when it purchased **Urchin** and subsequently, in 2006, released it as a free tool under the Google Analytics moniker. Now anyone who wanted to have access to first-class web analytics could do so for free. The number of customers using Google Analytics is half a million plus customers in the first six months. It is anticipated that Microsoft will soon follow Google and introduce a free web analytics tool.

Current Landscape and Challenges:

In the past few years, the Web has really “grown up” as a channel for most companies, and there is a deep demand for the web channel as the other channels (phone or retail, for example).

Web analytics = clickstream. Clickstream data is the source of all web decision making. Clickstream data is just a portion of web data. But there is lots of data and even more reports. And there is a statement that “The data is not telling me what I should do.”

The web analytics ecosystem is dominated by vendors trying to outdo each other by offering more and more features. Vendors set the analysis agenda.

Standard techniques such as **Customer-driven innovation (CDI)** have never taken deep roots in the world of web analytics. Most progress has been driven by **Possibility driven innovation (PDI)** represents “What else is possible for us to do with the data we capture? There is a lack of people and approaches that would enable web businesses to glean insights that result in action that enable strategic differentiation vs. their competitors. Universities and colleges are not teaching practical web analytics.

Web 2.0 and its associated technologies are increasingly becoming a part of the mainstream customer experience. This change is becoming a major disruptor for most

current web analytics approaches and vendors. It is even more important in the world of Web 2.0 that we accelerate the mindset shift and the strategy for implementing successful web analytics.

With the entry of **Google Analytics**, the market has simply exploded, because now anyone who wants to have access to data from their website can have it for free, and from a sophisticated tool to boot. Microsoft's anticipated free web analytics tool will only expand the options that practitioners have at their disposal. But access to the tool and data, although empowering, does little to ease the problems related to figuring out what your success metrics are and how to perform web analytics correctly. There is more data than ever available for a web analytics practitioner to tap into:

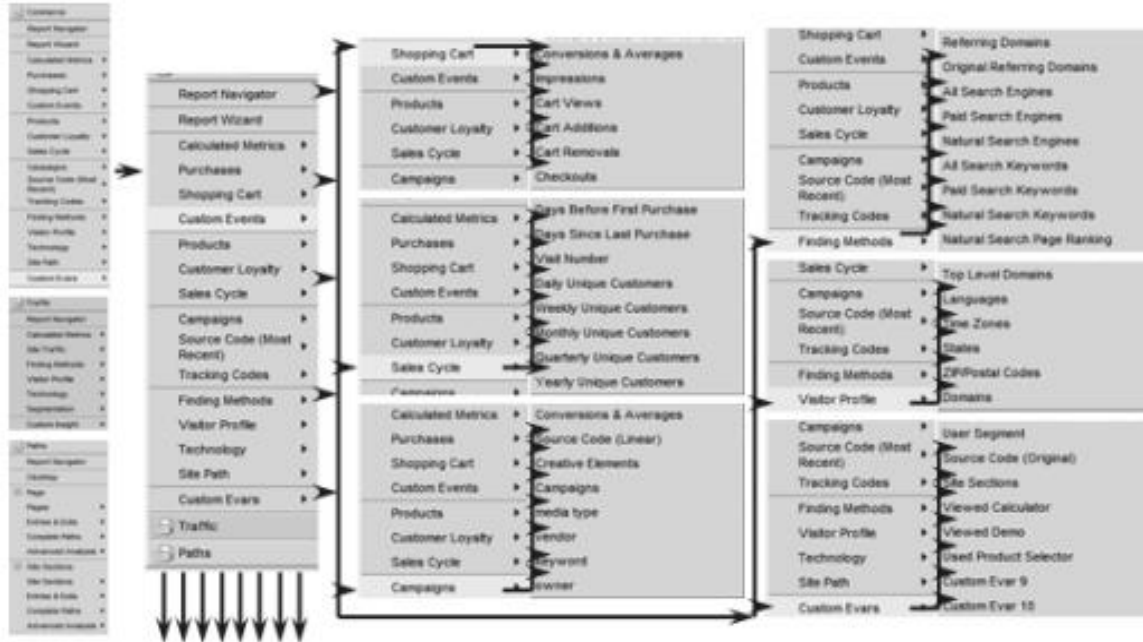
- Competitive intelligence to know not only what is going on at your site, but also (for a small fee) what is going on at a competitor's website.
- Qualitative data gives information about the effect of the web channel on the other channels (think Customer Relationship Management—CRM).

As web analytics has progressed to grow, an increasingly wide array of complex data has been made available. In almost every web analytics tool, it is now normal to see a couple hundred metrics at the click of a button.

This increasing amount of data provides an opportunity to become better at what we can analyze and act on. Companies in the web space spend millions of dollars on web analytics, campaigns and their websites, which are in turn chasing billions of dollars of online revenue. But, the number one challenge in surveys is the ability to measure accurately in order to make optimal decisions for those hundreds of millions of dollars companies spend. The reason this challenge persists is that most people go about solving it wrong.

Traditional Web Analytics Is Dead

Web analytics started its life with data sourced from web server logs, which primarily contain technical information and not business information. Because of this unique evolutionary path, the current crop of web analytics tools and the mindsets of customers are rooted in clickstream analysis. Figure 1.5 shows what web analytics has typically looked like.



Implementation of a web analytics tool takes just a few minutes, and instantly we have access to massive amounts of data, metrics, key performance indicators, stuff. There are practitioners and vendors and a well-established system of working and thinking in order to report on this data.

Traditional Web Analytics Metrics

1. Page Views

For an e-commerce website is it good or bad to have more page views per visitor? If you have frustrating navigation, and lots of page views—but no one will buy. If you have fantastic navigation, you'll have fewer page views—but maybe people decide faster that you don't have competitive prices and they leave anyway. Just from reporting page views, we may not find the case.

2. Hits

In the early days, hits tracked the requests that a server received to send data back. Hit can be translated as a page or content request. So more hits meant more content consumption, and it used to find visitors in the very early days. A typical page will cause 25 hits on a server. What can be really tracked using Hits? Requests for data to the server? Number of pages viewed? Number of visitors to the website?

3. Top Exit Pages

If you track the pages where more website visitors exit from the site, what does it tell you? That the pages are suboptimal? It could be that they are perfect pages where your customers find exactly what they are looking for and then leave. Consider me researching a Sony digital camera on Amazon.com. I find what I want, customer reviews, and I leave.

So do 99 percent of the people representing the traffic to that page. The exit rate doesn't tell you whether your content is good or bad.

4. **Website Engagement**

Engagement is often computed as sessions divided by unique visitors. If lots of people come again and again and have lots of sessions with your website, is it because they repeatedly can't find what they are looking for or because you have the most beautiful site in the world with perfect content.

5. **Visitor Screen Resolution**

Visitor screen resolution is a perfect example of a distracting metric that adds little value in any scenario. Every web analytics tool reports the monitor screen resolution of the website visitor and we have it in our daily reports, yet the metric rarely changes more than once every six months.

