

## Paper 180-26

**Know and Respond to your Web Customers**

Ben Zenick, Prowerk Consulting Group, LLC

**ABSTRACT**

Time and business needs have changed many things in the new economy, but one of the constants has been data. Businesses continue to have overwhelming amounts of data available to them to assist in making strategic business decisions. Over the years, another constant has been SAS® place as an industry leader in Data Warehousing and Decision Support Applications. By applying these time proven techniques they have begun to provide companies a very effective way for understanding the taxonomy of their web channels, as well as the behavior of customers browsing these various channels. This web channel, when utilized properly, can become one of the most important data sources of customer information and behavior. This data must be integrated into their strategic CRM vision in order to allow a business to quickly adapt and respond to market demands. . SAS WebHound Software™ provides you with “up to date information regarding your web traffic” and is a key tool within the SAS System for integrating Web channel data into your enterprise business intelligence systems. This paper will explain ways that you can utilize your web channels and their data as a competitive advantage in the rapidly growing e-Economy.

**INTRODUCTION**

As businesses move further toward creating e-Commerce channels, the needs of their customer base are rapidly changing as well. Instead of a focus on attracting customers to their stores, catalogs, etc. it is increasingly important for companies to attract their customers to their web channels and keep them coming back thus reducing churn and increasing the revenue per customer. As e-Commerce becomes the format that most customers use for purchases, what tools do companies have available to analyze these customers? Every transaction that takes place while surfing the web is captured in a log, which is essentially an ASCII flat file. These logs provide the information necessary to begin analyzing your web customers.

SAS has developed a solution called 'WebHound' to enable companies to do web log analysis and reporting. WebHound combines most of the SAS core technologies to create an environment that facilitates rapid reporting and analysis of web data. These technologies include, but are not limited to, the Base SAS transformation engine, SAS IntrNet®, and the SAS/Warehouse Administrator® software.

These components provide you with the real strength of WebHound: scalability and flexibility. Web customer analysis goes beyond analyzing web logs. Every business has other customer touch points, and views their customers differently. WebHound provides you with the capability to view your customers based on your business rules.

What information are people interested in analyzing? Some of the questions WebHound helps answer include, but are not limited to:

- Who is visiting my web site?
- Where are they going?
- What are they browsing for?
- How long were they there?
- Did they purchase? Why not?

Many other questions about the behavior of your web customers can be answered.

In order to really use your web channels, and the resulting information as a competitive weapon, SAS has empowered you with WebHound to facilitate the process of answering business questions about your channels and how customers interact with them. Before you can analyze your customers, you really need to understand how to obtain and keep them. There are many techniques available to assist you in creating an ongoing relationship with your customers, as well as gain an understanding as to what the most profitable customer profile is.

**CUSTOMER RELATIONSHIP MANAGEMENT (CRM)**

Why CRM? Customer Relationship Management is a comprehensive approach which provides seamless integration of every area of business that touches the customer - namely marketing, sales, customer service and field support-through the integration of people, process and technology, taking advantage of the revolutionary impact of the Internet. CRM creates a mutually beneficial relationship with your customers. One of the goals of creating a CRM environment is being able to create an intimate environment for your customers. Every customer is different and should be treated in a manner that identifies their individual needs. They must feel a sense of comfort or intimacy with you and your web channels in order to complete a purchase. Customers tend to lock into a “safe harbor”. Your sites goal should be to create such a high comfort level with your customers that intimacy is a given. This intimacy can only be achieved by using every tool available to understand your customers.

One of these tools is your web site. It is nothing more than another touch point where you can interact with your customers. However, this interaction also creates more data that companies must be able to manage and derive strategic business decisions from. So, it becomes imperative to understand what information is available, and how that information can be combined with other touch points. Some of these other touch points include:

- Call Centers
- Catalogs and Brochures
- Direct Mailings
- Promotional Campaigns
- Direct to Stores

All this information will not only give you the capability to understand the taxonomy of your web site and the web customers behavior, but the customers behavior in relation to your entire company.

When managing a CRM environment you must be able to connect all of the customer touch points into one centralized environment. Each of these touch points have the potential to create a great deal of data. WebHound creates a solution that is flexible and scalable enough to allow companies to incorporate web data with data from other touch points throughout your organization to create one customer centric view. The customer

centric views allow organizations to begin to get a feel for what their customers needs are, and respond to them in a timely manner.

## WEB DATA

Web Servers can record every click that a user makes when browsing your web pages. This results in extremely large amounts of data, in the form of web logs, which must be analyzed. A web log is a text file that records every request that a client makes to your web server. This information creates the so-called "click stream". By analyzing these clicks, you can begin to answer many questions in regards to how customers are interacting with your web site

To imagine the possibilities this data can create, we will look at an example of a "Brick and Mortar" business that we are all familiar with, a retail store. Every time a customer visits a retail store, nothing is known about the visit until something is actually purchased. You are not followed through the store with a record being kept of each item you think about purchasing. This remains an unknown to anyone but you. The only thing about the visit that is recorded is the purchase itself. In the age of "Click and Mortar", the name of the game changes, since these stores now have the capability of recording the entire scope of a user experience, not just the purchase. This information can create an environment where the experience can be enhanced for each individual user by generating knowledge from their click stream.

By itself this click stream data can be somewhat limited. WebHound provides you with the capability to create a "click stream data mart" that is scalable and open. This gives you the opportunity to create a solution capable of integrating other mission critical legacy applications, enterprise business packages, CRM solutions, and other internal systems.

A study conducted by the Hurwitz Group, can assist us in uncovering some of the potential gold at the end of the rainbow that click stream data can provide. The study can also assist us in beginning to understand "Why we should collect the clicks". The study breaks the value we can gain from web log analysis down into five different response groupings. They include the following:

1. **Behavioral Analysis.** Using the click stream to begin profiling an individual user's behavior
2. **One-to-One Marketing.** Creating user profiles to be used to generate Web-page content dynamically.
3. **Path to purchase.** Creating the path of least resistance, to reduce the number of clicks a user must make before buying.
4. **Product affinities.** A customer's path through your site might unveil product affinities not normally seen.
5. **Usability.** Click stream analysis can show usability problems in your web site. Do you have web pages that take too many clicks for users to get to?

All of these issues can be resolved by having an effective way of analyzing the activity that takes place through a variety of web channels you may have implemented at your company. This analysis can also include things as simple as:

What are the most popular paths users are taking through my web site?  
How often are certain pages being accessed?

Before we can begin answering the questions we have about our web site, and the customer's interaction, we must first understand web logs, and the information available in them. In order to do this, an understanding is needed of some of the potential problems encountered when attempting to analyze web logs.

As discussed earlier, Web Servers write each click a user makes to a web log. This poses a potential problem for an out of the box web log analysis solution. As shown in the table below, there are many different types of web servers.

| Server Types  | Characteristics  |
|---------------|--|
| Apache        | Highly Customizable<br>Emulate any form of web logs except ELF<br>Save logs in /logs/access_logs |
| Microsoft IIS | Extended Log File Format<br>ODBC Logging<br>Save logs with extension exmmdd.log                  |
| Netscape      | Usually CLF2<br>No Standard Format   |

**Table 1: Web Servers**

The above web servers, and the others that are available, each create logs with differing characteristics. Some of these formats can be seen in Table 2 below.

| Web Log Type | Characteristics   |
|--------------|---|
| CLF          | Contains only basic set of web log information<br>No Referrer or User Agent |
| Combined CLF | CLF + 2<br>Contains CLF fields plus Referrer and User Agent                 |
| ELF          | Includes Referrer and User Agent<br>Can contain Cookies                     |
| Netscape     | Proprietary Format  |

**Table 2: Web Log Formats**

This list does not really shed light on the difficulty in reading and analyzing logs. Not only can logs be produced from a variety of different servers, but each of these different web logs (which don't even make up the entire population of logs) is also customizable.

Part of creating a flexible solution for web log analysis is being able to handle all the different scenarios we see. The solution must be capable of reading data from different web servers in different log formats. WebHound provides an easy to use interface for dynamically reading different types of Web Logs into a "click stream data mart".

As shown in Figure 1 below, web log data can be extremely cryptic information. Thus, it is necessary to have tools flexible enough to manipulate the data accessible in the logs.

```
210.94.6.103 - - [13/Jun/2000:00:15:01 -0700]
"GET /images/interactive/f_deltoidr2_ov.jpg HTTP/1.1" 200
385
"http://www.myprimetime.com/health/tool/content/virtual_workout/index.shtml"
"Mozilla/4.0 (compatible; MSIE 5.01; Windows 98)"
"210.94.6.103-96087978839" 0 0
```

### Figure 1: Web Logs

At first glance it may appear this data does not provide you with a lot of information to help you understand your web customers. In reality, there is a lot of valuable content contained in the text stream shown above. This one record contains information such as:

- Who is the user-IP Address
- When did they click-datetime
- What action took place-GET/PUT
- What page was requested-URL
- Where did they come from-Referrer
- Action Status Codes-HTTP Status
- Browser types and platforms

Other information can also be obtained from one click, depending on the type of web logs being analyzed, as well as how customized they are.

There must be a lot of thought into designing how you will be extracting the web log data to create your mart. Since web servers track every click that is made to your web channel, there is no direct correlation between the amount of data in a web log, and the number of customers. Large volumes of data don't necessarily mean many visitors. It could mean that the data you are analyzing has been modeled to keep every image, string, etc, downloaded.

These large volumes of data mean that a good design has to be put in place, or analysis will be virtually impossible. The data must be summarized up to a level that can be used for analysis. This level will be based on your business rules. The web log analysis solution that your company chooses must have the capability that allows users to change levels of summary as well as how much detail, and what detail should be kept in the mart.

Many companies host web channels as part of their business. In order to get paid, they must accurately determine how many hits each of the unique servers have. It is not uncommon for a host to have to report on several hundred different web channels at a detailed level. Based on business rules and SLA's this level of detail could be represented by every object that gets downloaded to just a single page. The needs are usually not consistent since no company has the same business rules especially when it comes to what a page count is. WebHound can be used in situations like this to ensure the integrity of the data from each web server, by having the capability of implementing different business rules, at different aggregation levels. The solution gives you the capability to get accurate page counts for the different web servers thus giving you the capability to charge customers for your services.

### CONCLUSION - WHY WEBHOUND?

By using a solution like WebHound, you are armed with all the power and flexibility of SAS. Since SAS has a Multi-Engine Architecture (MEA), it is very easy to integrate other data sources with your web logs. MEA gives users the capability to read data in any format, from web logs (text files) to DBMS'. This is important because reading web logs is only one part of analyzing your web customer. All this data needs to be integrated into the bigger picture by creating a customer centric data warehouse. This is most effectively accomplished by taking a look at all the touch points of your company and integrating them with other systems. With SAS' Multi Vendor Architecture (MVA), it is irrelevant what platform the data exists on. Using these techniques along with other components of SAS' software suite, web logs become yet another data source for making strategic

business decisions about your customers.

### REFERENCES

DM Review Magazine – Web articles  
 CIO Magazine  
 WebHound Custom Course Notes – written by Ben Zenick and Gary Jackson (SAS Institute Inc.)

### ACKNOWLEDGMENTS

Thanks to Joseph Costanzo (Prowerk Consulting Group, LLC) and Robert Fecht (Prowerk Consulting Group, LLC) for input and review of this paper.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at: <mailto:ben.zenick@prowerk.com>