

Paper 016-31

## SAS® Web Analytics 5.2: What's New?

Jonathan Polito, SAS Institute Inc., Cary, NC

### ABSTRACT

This presentation describes and demonstrates some of the new features of SAS Web Analytics 5.2. SAS Web Analytics collects, analyzes, predicts, and reports Web-site visitor behavior. This information helps businesses better understand the effectiveness of their online (and, potentially, offline) initiatives, and the information optimizes their Web site and business practices. These new features are discussed:

- Visitor drop-off; funnel and path analyses
- Report creation, navigation, and linking
- Decision support analytics and integration of additional log data
- Server performance improvements
- Page tagging
- Use of powerful expressions and synthetic columns in reports

### INTRODUCTION

More and more information and more demand for attention are being streamed into the world. The competition to attract and retain customers is fiercer than ever. Globalization, the regulatory environment, technology, and more demanding customers make it imperative to be competitive and to distinguish your business from everyone else.

A successful company uses a customer-centric approach; this is possible only by understanding all customer touch points. Many companies have built and invested in large data warehouse projects and software to help achieve this end. However, these companies have left off what is now becoming the fastest-growing customer touch point of all, the Web. By being able to understand this customer touch point with all the other customer channels gives you a complete view of the customer. This allows you to have a truer understanding and analysis of your visitors and allows you to improve conversion and customer retention rates.

Most businesses or organizations have a Web presence now. Furthermore, the Web is capturing ever-larger monetary amounts and interaction opportunities. Many channels—from email, mass mailings, and TV ads to Web search engines—drive visitors to your Web site. By capturing your Web site's click stream, you have a convenient integration point for the analysis of multiple data channels such as call centers, products, campaigns, e-commerce, and offline commerce and order completion.

SAS Web Analytics 5.2 analyzes and reports on your multichannel data sources that are centered on your click stream. This paper provides motivation for this analysis, our philosophy regarding the analysis, and new SAS Web Analytics features that make this analysis possible.

This paper uses the term *visitor* to refer to the originator of the click-event on your Web site. Alternatively, you can substitute *customer* or *user* for visitor. Visitor, while perhaps less specific, seems more accurate because all customers for your Web site are visitors, but all visitors are not customers. Moreover, a goal of Web analytics is to help you determine how to turn visitors into customers. This paper defines a *visit* as the visitor entering your site, viewing one or more pages and leaving.

Also, this paper uses the term *organization* to refer to the Web site owner. Again, organization is more accurate than the term *business*, for the general case, because all businesses are organizations, but not all organizations are businesses.

### BACKGROUND

SAS Web Analytics 5.2 uses a number of novel, as well as standard, methods to assist you in understanding your visitors' behavior. SAS Web Analytics 5.2 provides a highly customizable extract-transform-load (ETL) process for reading your Web logs, associating the stateless individual clickstream records with a session and then analyzing and summarizing the session and clickstream traffic. This data resides in an optimized Web data mart for analytical reporting. A mid-tier Java Servlet and JSP-based Web application sits on the server data tier that hosts the Web data mart. SAS Web Analytics 5.2 provides two Web-based applications:

- the SAS Web Analytics Administrator Web Application for administrating the analytical ETL process, creating and customizing the Web analytical reports, and managing your Web data marts
- the SAS Web Analytics Viewer Web Application for viewing the reports and visualizing the analytics.

With basic traffic analysis, you can understand how your visitors enter your Web site and how they leave. You can identify what pages they land on and with what frequency. If the visitor comes from a search engine, you can identify what search terms they used. You can determine how many unique and repeat visitors are coming to your Web site. When a visitor comes to your Web site, you can discover site problems by examining the error codes and associated pages. By knowing what browsers and platforms your visitors use, you can ensure that your site presents well. Many groups can use basic traffic reports, for example, those groups responsible for editorial content, usability, and making strategic decisions such as what browsers to support.

Going beyond traffic analysis, SAS Web Analytics provides *decision support*. This is the process of gathering and analyzing data from a wide variety of sources in order to make decisions based on the data. SAS Web Analytics 5.2 provides a rich decision-support system consisting of analytical dashboards, scorecards, and segmentation.

As a first step, you can create an analytical dashboard containing the key performance indicators (KPIs) that you select and categorize for your organization. The dashboard gives you an overview of how your KPIs are performing and trending according to the business directions that you set. To improve a KPI, you can create an analytical scorecard that contains the KPI as the scorecard's target. Then, select all the potential or possibly interesting drivers for that target. The analytical scorecard process will then use a stepwise regression that determines which drivers are associated with the target. After determining the associated drivers, econometric forecasting identifies trends for each driver and the target. In addition, the scorecard process will use the model derived from the associated drivers on the target to create a series of calculations that are displayed in the goal-seeking table. You can use the goal-seeking table to answer "what if" questions. Goal seeking shows you by how much you must increase a driver in order to get (for example) a 5% increase in the target.

Segmentation analysis defines a set of homogeneous groups or characteristics (segments) for visitors who have performed a specific action that is your target. In this analysis, you choose (a) a target (or response) variable and (b) a set of independent variables that are potentially predictive. SAS Web Analytics 5.2 then uses a decision-tree modeling technique, based on the ARBORETUM procedure, to predict the segments or groups of the independent variables that are characteristic of the target.

You can explore visitor behavior through the use of *funnel analysis* or *path analysis*. In many Web sites, areas of the site require a visitor to progress through a set of sequential pages. For this case, you define funnels that specify the steps through your Web site that you are interested in analyzing. The SAS Web Analytics Funnel Analysis then displays the summary statistics for the visitors through the funnel and the associated drop-off. For example, if you had a registration process that consisted of a number of pages, the funnel analysis can alert you to excessive drop-off for a certain transition and areas to optimize for retention.

To determine what visitors actually view after a starting page, prior to an ending page, or between a starting and ending page, use path analysis. Path analysis helps you visualize how your visitors are moving through your site, where they are entering and leaving your site, and highlight poorly designed paths through your site.

This brief overview has concentrated on Web data specifically. Throughout the analytical ETL processing, there are points to integrate other data channels into your Web mart and provide a more systemic analysis. For instance, you might have an online mortgage application process that you monitor. In addition to the usual analysis involving Web data, you can integrate an offline data source such as completed and submitted paper applications. In the United Kingdom after an online application is submitted, regulations require the provider to send a hard copy of the mortgage and necessary contracts through the postal service. The customer must then complete the paper contracts and mail them back to the provider. A simple Web site mortgage application conversion analysis would leave out this important step. With SAS Web Analytics 5.2, you can include this in your analysis.

## CALENDAR-DRIVEN REPORT NAVIGATION

SAS Web Analytics 5.2 includes a convenient and intuitive calendar control for powering your reports and visualizing the available data. Once you select a report, you can navigate across available dates and date ranges, simply by clicking on the calendar. The calendar is report-aware and will boldface only days for which there is data as shown in Figure 1.



Figure 1. Calendar Control

You can instantly see gaps in your data.

Select a date range by selecting the start date and the end date. If you want to select an entire week, use the week selection control  button. You can also select a date range simply and quickly by using the “M,” “Q,” or “Y” quick selections for month, quarter, and year, respectively. Moreover, if you want to navigate to the start or end of your data’s range, use the “S” or “E” control. If you need help using the calendar, there is calendar-specific help that is available now by clicking the Help  icon.

The calendar-driven report navigation is available for all report classes. That means that there is a report-aware calendar for any report in *Traffic*, *Scorecard*, *Dashboard*, *Funnel*, *Path Analysis*, or *Segmentation*. Also, different groups have slightly different calendar controls. For instance, scorecards and dashboards are day-based; thus, there is only day selection. Likewise, path analysis and funnel analysis can be across noncontiguous days; thus, you can select any number of days individually as shown in Figure 2.



Figure 2. Calendar for Funnel Analysis

## REPORT AND ANALYSIS LINKING

SAS Web Analytics 5.2 supports report and analysis linking across a variety of reports out-of-the-box. You also can configure your own report-linking through the Web-based administrator interface.

If you have a key performance indicator (KPI) in a dashboard that is also a target in an Analytical Scorecard, then there is, by default, a link from the dashboard KPI to the scorecard. Likewise, if a scorecard contains a metric that is a target in another scorecard, there is a link from the metric in the first scorecard to the scorecard with that metric as a target. Also, you can manually set a link from a scorecard metric to a traffic report.

For analytical dashboards, you can link a dashboard KPI to another dashboard as well as add a link to a traffic report. Conversely, you can link a traffic report to another traffic report.

### SAMPLE WORK FLOW

Your dashboard signals that your KPI, *Found a Mortgage*, is trending in a negative business direction. You look at the trend graph and notice that there is some *seasonality*, since the KPI peaks early in the week and decreases through the rest of the week. You notice that you are on the Friday downward trend.

This KPI, *Found a Mortgage*, is also a target in Analytical Scorecard. You follow the link from the dashboard to the analytical scorecard, *Mortgage Selections*, containing the target KPI (*Found a Mortgage*).

From the Analytical Scorecard, you can see that one (or many) of the associative metrics has a negative trend and is outside the predicted range. Alternatively, the associative metrics might be in a steady business state. This indicates that the business state is steady in the longer term (30 days). However, the business state trend might be negative in the shorter term (7 days).

Perhaps, you want to focus on the associative metric, *Repeat Visitor*. From this associative metric, follow the link to the *Site Metrics* traffic. On this report, select a full week of reporting and examine the trends between unique and repeat visitors.

## EXCEL INTEGRATION

Many business users are comfortable with Microsoft Excel and want its layout and multimedia capabilities. Each report in SAS Web Analytics 5.2 containing a table now can generate a URL that is suitable for using in Microsoft Excel's *Web Query* functionality. To create an Excel report that will be driven by live SAS Web Analytics data, you need to first locate the unique URL to serve up the data:

```
http://itvdev6:9090/d0wa52/Wax?tag=document&export=true&documents=100&document=8a7ec5e0_ae0b130a_00ffeef1_52b0464b&fromdate=2005-1-30&todate=2005-1-30
```

The SAS Web Analytics Web application generates this URL when you select the *Generate URL for Export*  icon on a SAS Web Analytics table as shown in Figure 3.

	Requested File ▲ ▼	Entry Page Count ▲ ▼	Percent Entry Page Count ▲ ▼
1	<a href="/store/product.aspx?catalog_name=books_and_audio">/store/product.aspx?catalog_name=books and audio</a>	155	19.72
2	<a href="/default.aspx">/default.aspx</a>	70	8.91
3	<a href="/store/default.aspx">/store/default.aspx</a>	31	3.94
4	<a href="/library/library.aspx?libraryid=11338">/library/library.aspx?libraryid=11338</a>	27	3.44
5	<a href="/activate.aspx">/activate.aspx</a>	21	2.67
6	<a href="/businessstopichome.aspx">/businessstopichome.aspx</a>	21	2.67
7	<a href="/keycode.aspx">/keycode.aspx</a>	19	2.42
8	<a href="/registration/subscribe.aspx">/registration/subscribe.aspx</a>	16	2.04
9	<a href="/library/library.aspx?libraryid=11368">/library/library.aspx?libraryid=11368</a>	15	1.91
10	<a href="/seminars/seminar.aspx?libraryid=11240">/seminars/seminar.aspx?libraryid=11240</a>	8	1.02
11	<a href="/seminars/seminar.aspx?libraryid=4896">/seminars/seminar.aspx?libraryid=4896</a>	8	1.02

Figure 3. Entry Pages Table

After you click the *Generate URL for Export* icon, the application gives you a dialog box with the URL (Figure 4) that you can copy to your clipboard.

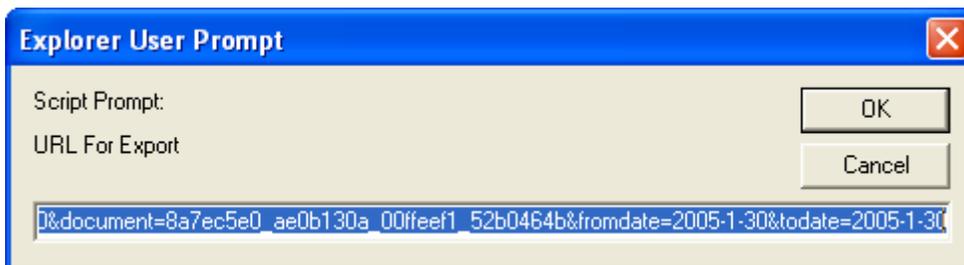


Figure 4. URL for Live Data Query

In Excel, select Data -> Import External Data -> New Web Query (Figure 5).

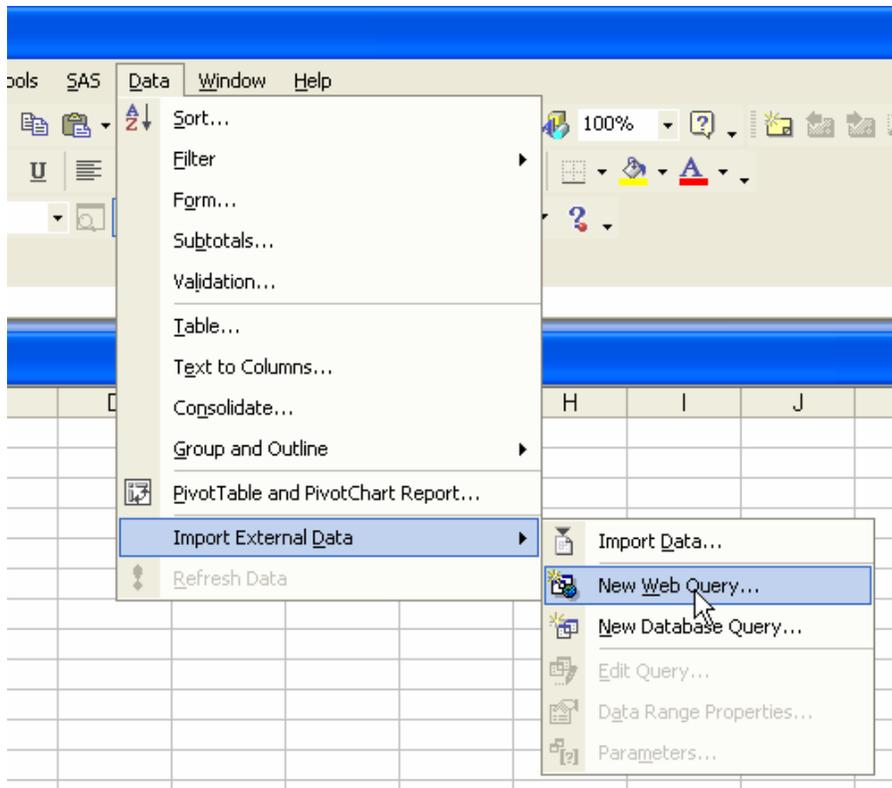


Figure 5. Excel New Web Query

Next, paste the URL into the address box and select **Go**. Under "**Options...**" in the *New Web Query* dialog, turn on Full HTML Formatting. Now, select **Import** and choose where you want to put the data in the worksheet (Figure 6).

 The image shows an Excel spreadsheet with the following data:
 

	A	B	C	D
1		Entry Pages (1/30/05) (1 to 50 of 283)		
2		Entry Page Count greater than 0		
3		Requested File	Entry Page Count	Percent Entry Page Count
4	1	/store/product.aspx?catalog_name=books_and_audio	155	19.72
5	2	/default.aspx	70	8.91
6	3	/store/default.aspx	31	3.94
7	4	/library/library.aspx?libraryid=11338	27	3.44
8	5	/activate.aspx	21	2.67
9	6	/businesstophome.aspx	21	2.67
10	7	/keycode.aspx	19	2.42
11	8	/registration/subscribe.aspx	16	2.04
12	9	/library/library.aspx?libraryid=11368	15	1.91
13	10	/seminars/seminar.aspx?libraryid=11240	8	1.02
14	11	/seminars/seminar.aspx?libraryid=4896	8	1.02
15	12	/learningseries/learningseries.aspx	7	0.89
16	13	/library/default.aspx	7	0.89
17	14	/library/library.aspx?libraryid=9178	7	0.89
18	15	/store/_additem.asp?catalog_name=university_catalog	7	0.89
19	16	/businesstopicindustryindex.aspx	6	0.76
20	17	/library/library.aspx?libraryid=11367	6	0.76
21	18	/rss/newsfeed.aspx?libraryid=10875	6	0.76
22	19	/seminars/seminar.aspx?libraryid=10943	6	0.76
23	20	/solutions/infoacklist.aspx	6	0.76

Figure 6. Web Analytics Data in Excel

With this live query to SAS Web Analytics, you can do additional computations in Excel or produce your own Excel-type reports and graphs. Again, an important point is that this is a query to live data and when you open your report, you always pull the most recent data.

## FUNNEL ANALYSIS

SAS Web Analytics 5.2 provides a well-defined funnel definition and analysis. A funnel is defined as a path through your Web site that visitors may traverse. The funnel is composed of an ordered sequence of URLs, or levels, that define the path. Furthermore, you may constrain a given level to be adjacent to the previous level. This means that the URL on the subsequent level must follow directly from the URL on the previous level, without any intervening URLs.

SAS Web Analytics 5.2 adds several improvements:

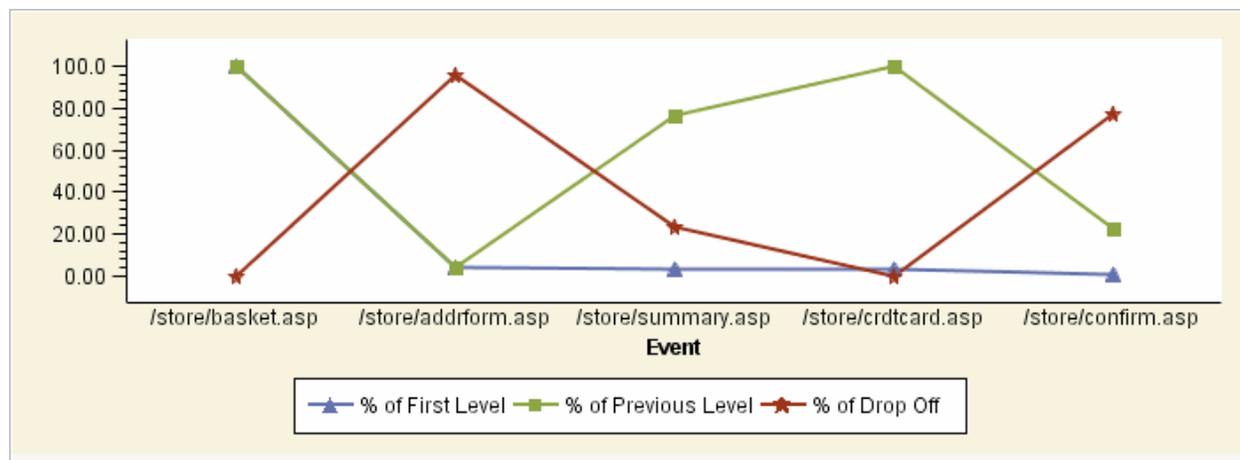
- An improved user interface with search and filtering capabilities
- Static funnel support (1, 7, and 30 days)
- Stricter adjacency definition

With funnel analysis, you can investigate how visitors move through the paths on your site. You can determine if their behavioral intent matches your site's intent. You still need to understand your organization and your site itself. You need to know what your organizational intents are on the site, whether they drive the visitor to a product purchase, to a paper download, to an FAQ (frequently asked question), to a registration, or to specific content. Then, given the organizational intent, you reduce it to a path or set of steps that you want the visitor to progress through on your site.

For instance, you can look at visitor behavior for the past four weeks for product purchases from your Web site by defining a funnel that consists of the non-adjacent pages:

1. /store/basket.asp
2. /store/addrform.asp
3. /store/summary.asp
4. /store/crdtcard.asp
5. /store/confirm.asp

Next, look at how visitors transitioned through each level (Figure 7).



Event	Total Visits	Path Visits	% of First Level	% of Previous Level	% of Drop Off
/store/basket.asp	425	425	100.0	100.0	0.00
/store/addrform.asp	21	17	4.00	4.00	96.00
/store/summary.asp	16	13	3.06	76.47	23.53
/store/crdtcard.asp	15	13	3.06	100.0	0.00
/store/confirm.asp	4	3	0.71	23.08	76.92

Figure 7. Product Purchases

Observe that you had a 96% drop-off between the shopping cart (/store/basket.asp) and the address form (/store/addrform.asp). If a visitor made it to the address form, then you had about a 24% drop-off going to the order summary and then no drop-off going to the credit card form. Unfortunately, you had an almost 77% drop-off going to the confirmation page.

This analysis gives you a number of opportunities for further investigation and ultimate improvement. Probably, the first areas that you would review are the visitor behavior from the shopping cart and the visitor behavior from the credit card form. You can investigate this behavior with the *path analysis* feature in SAS Web Analytics; this feature can help you understand what the visitors are doing at these points. Path analysis is discussed in the following section.

#### **FUNNEL ADJACENCY**

For SAS Web Analytics 5.2, funnel analysis will give a visitor credit for traversing a funnel once per session and will select only the largest funnel for a session. You can consider a funnel in pair-wise segments. Thus between any two levels, with the first level having a value of A and the second level having a value of B, we have two cases:

- AX\*B (non-adjacent case)
- AB (adjacent case).

For instance, the following paths would be part of the funnel defined by:

- AB
- ACB
- ACDEFB.

For the adjacent case, only the path "AB" is part of the funnel.

Thus, the following paths are part of the funnel if you define a funnel as AX\*BCDX\*E:

- ABCDE
- ABBBCDE
- AQRSBCDYZE.

#### **FUNNEL USER INTERFACE**

Some sites have more than 10,000 pages. Most people would not want to have to manually search through those pages to find the one they are looking for or be presented with a list of every single page. SAS Web Analytics 5.2 includes a searching and filtering capability to make page selection much easier. When you select a page for a level in the funnel, SAS Web Analytics presents you with a list of all the available pages (Figure 8). You can filter or search by entering in a fragment of the page you are looking for to effectively reduce the pages you must examine for selection (Figure 9).

**Interactive Funnel Form**

**Select Funnel**

Level	Url	Adjacent
1	<input type="text"/>	
2	<input type="text"/>	<input type="checkbox"/>
3	<input type="text"/>	<input type="checkbox"/>
4	<input type="text"/>	<input type="checkbox"/>
5	<input type="text"/>	<input type="checkbox"/>
6	<input type="text"/>	<input type="checkbox"/>
7	<input type="text"/>	<input type="checkbox"/>
Is	<input type="text"/>	

- /about/contact.aspx
- /about/default.aspx?libraryid=2964
- /about/default.aspx?libraryid=3017
- /about/default.aspx?libraryid=3018
- /about/default.aspx?libraryid=3019
- /about/default.aspx?libraryid=8885
- /about/default.aspx?libraryid=8915
- /about/default.aspx?libraryid=8930
- /about/howtouse.aspx
- /about/pressreleases.aspx
- /about/pressreleases.aspx

Figure 8. Funnel Page Selection

Level	Url
1	<input type="text"/>
2	<input type="text"/>
3	<input type="text"/>
4	<input type="text"/>
5	<input type="text"/>
6	<input type="text"/>
7	<input type="text"/>
Is	<input type="text"/>

- /store/\_additem.aspx?catalog\_name=video
- /store/catalog.aspx?catalog\_name=video
- /store/category.aspx?catalog\_name=video
- /store/category.aspx?catalog\_name=video&category\_name=video

Figure 9. Funnel Page Filtering

### STATIC FUNNEL SUPPORT

Interactive funnel analysis is useful for ad hoc investigation and discovery of visitor site behavior. With interactive funnel analysis, you can vary the pages within the path, the adjacency of each level, and the number of days considered. However, once you have settled on a set of your favorite funnels, you can eliminate the cost of the interactive funnel analysis stored process with static funnels.

New in SAS Web Analytics 5.2 is the ability to define a static funnel analysis that will be run during ETL processing time for predefined historical periods of 1, 7, and 30 days. When you define or modify a funnel, you have the option of adding that funnel definition to the nightly ETL processing window. This is convenient for funnels that you have already identified and that you look at regularly. It is also convenient for large funnels or sites with high traffic volumes since you do not have to wait for the interactive process to finish.

### PATH ANALYSIS

When you want to investigate visitor site behavior based on starting and ending points on your site, you can use the path analysis in SAS Web Analytics 5.2. With this analysis, you select a starting page (or pages), ending page, or both. Using path analysis, you can learn where visitors go once they enter your site, either directly or through a referrer. Alternatively, you can learn how they got to a certain page.

From the funnel analysis example described above, you can learn from path analysis what happened to the visitors between the credit card form and the confirmation page. Then you perhaps can improve the drop-off rate (Figure 10). In this situation, you have a clear case of simple abandonment. After the credit card form, your visitors are just leaving the purchase process. Perhaps this indicates that visitors are only looking for some specific information on that page and have no intention of confirming their decision, or maybe they lose interest once they learn further details like high shipping charges. Maybe visitors have difficulty determining how to confirm their purchases. In this case, note that the path analysis entry page is based on the total visits to that page, rather than the reduced visits in the funnel analysis as a result of the funnel traversal.



Figure 10. Path Analysis

SAS Web Analytics 5.2 features a redesign of its path analysis. At the base of this redesign is a new PATH procedure with performance improvements as well as calculation of additional statistics. The SAS Enterprise Miner PATH procedure can now operate off of our pathing data set directly rather than having to create an intermediary data set generated from the DMDB procedure. This eliminates one pass of the data in the process. Usually, the pathing data set is the largest data set in the Web mart so this performance improvement helps with both space and time.

SAS Web Analytics utilizes new capabilities of PROC PATH in order to compute statistics on the interior nodes of the path such as Visits, %Visits, %Dropoff, and %Root. We utilize the new funnel feature of PROC PATH in order to create a funnel data set that is then summarized along with the primary path data set.

With path analysis, you can choose whether or not you want the noninteractive path analyses to run in the analytical ETL process. If you are backloading multiple days, you might want to wait until the backload is completed before running the path analyses, because it is a computationally expensive operation and you need only 1-, 7-, and 30-day reports for the pre-computed path analyses. You can always run interactive path analysis for any time interval if you need an ad hoc interval or path.

The path analysis visualization also has been reworked with better graphics and an improved statistical display. The nomenclature is now consistent across the interactive and static path analyses. The visualization utilizes the new SAS SILK graph components (such as the OrgChart) that allow better layout, annotation, and manipulation.

SAS Web Analytics 5.2 now gives you the additional flexibility of specifying the minimum and maximum sequence lengths for the interactive path analysis. This added parameterization improves the selection capabilities of the path analysis because you can limit the paths under investigation to meaningful lengths.

### **ANALYTICAL ETL ENHANCEMENTS AND PERFORMANCE IMPROVEMENTS**

Two ETL processes comprise SAS Web Analytics ETL. The first ETL process extracts the data from the Web log or logs, applies filtering and mapping functions, and effectively turns the stateless discrete records in the Web log(s) into cohesive sessions that are associated with a visitor. This is called *sessionizing*.

After sessionizing, an analytical ETL process (WAETL) runs; it further processes the data and augments it with fields that aid in the analysis and production of summary tables. The summary tables are report- and analysis-ready and help SAS Web Analytics maintain reasonable performance even when faced with the voluminous amount of data produced by Web traffic.

SAS Web Analytics 5.2 contains several major enhancements in its ETL processing. Additionally, we've increased the internal limits for a number of WAETL parameters such as the number of name-value pairs in the referrer query string as well as made adaptive some internal array limits.

#### **STREAMLINING THE WAETL LOAD**

In normal production mode, you process a single day of Web logs during the ETL batch window. However, often you will find yourself having to do back loading of data. This is especially true when you are first setting up a site for analysis. In this case, you will process 30, 60, 90 or more days worth of logs while you are constructing your Web data mart. SAS Web Analytics 5.2 addresses this with a number of new features that streamline the analytical ETL.

Primary among these new features is the ability to exclude path analysis during the ETL processing. Path analysis is one of the most costly processes within SAS Web Analytics and because the predefined path analyses are only based on the past 1, 7 and 30 days worth of data, you can now hold off the analysis until the last day of your processing during back loading.

#### **SUMMARY ENHANCEMENTS**

You can now also select to only include certain summaries as well as particular decision support reports during WAETL time. Sometimes this will help your performance, but more often you will use this new feature to integrate external data sources or use while you are prototyping a new summary or decision support analysis.

SAS Web Analytics 5.2 provides a capability to run a summary in test mode or production mode. When running in test mode you can select to either update or overwrite the summary output data set. For instance, in test mode you generally want to replace any summary tables that may exist. In production mode, you will append to existing summary tables, and then re-summarize for the class variables. Again, this is useful during prototyping.

You can use the summary test mode to experiment and make sure that the summarization that you design is correct before adding it to the batch ETL process. Furthermore, by running the summary in test mode, you have an advanced view of what the summary output table will look like. Thus, you can then design a report based on the new summary in the SAS Web Analytics Report Administrator and once you finalize your summary design, you can add it to the production ETL process. Then when you are satisfied that the report is correct, you can, with a simple toggle, make the report available in the SAS Web Analytics Viewer immediately.

An additional new feature that helps you create your summaries is the copy summary feature. When you copy a summary, you will have the option of modifying any existing options before creating the new summary. Obviously, this helps when you have a number of similar summaries that you must define.

#### **PERFORMANCE IMPROVEMENTS**

The SAS Web Analytics stored processes now use the compress option by default which can drastically reduce the amount of data stored because Web data often contains fields that aren't always filled (like the referrer field which rarely requires the maximum capacity). Often we attain compression ratios of 95% for Web data.

Throughout the SAS Web Analytics ETL process, we paid considerable attention to improving the performance. We refined a number of algorithms, eliminated redundant steps, and structured code to run in a more optimal manner. Between the past two releases, we see improvements like those shown in Table 1. We made these timings using 31 days of our favorite Web test data.

	5.1	5.2	Percent Difference
Real Time	35:09:00	18:11.66	-48.2
User CPU Time	22:48.15	13:33.96	-40.5
System CPU Time	22:48.15	13:33.96	-39.1
Memory (KB)	84,246,807	61,408,335	-27.1

**Table 1: Performance Improvements**

You see these improvements primarily as a result of the path analysis optimizations. When path analysis is run every day (for instance, in the non-back loading scenario), the performance improvement for SAS Web Analytics 5.2 is roughly 10% better than the 5.1 version.

### DECISION SUPPORT

Across the decision support features, SAS Web Analytics 5.2 allows you to determine whether or not to run the process at ETL time. Again, this helps when you are designing new decision support routines or when integrating an external data source that is unavailable to WAETL.

This control is also very useful when you have external data sources that you must load independently of the analytical ETL process. For instance, you might have to integrate your summarizations produced during analytical ETL with an external data source consisting of updated e-commerce purchases. After the data integration is completed, you can run the specific decision-support routines based on that data to analyze the combined Web and external data sources. Another new feature that makes this decision support chaining possible is the convenience of using the dates from a previous WAETL step rather than having to explicitly provide them.

### TAGGING

The tagging of pages involves inserting a small (usually JavaScript-based) snippet of code in all the pages of your Web site. The snippet of code then adds information in the form of request parameters to a request for a small image file, usually a 1x1 GIF. This request can go back to a dedicated tagging application server or simply go back to the originating Web server and generate a detailed log record. Most often, the page tag is coupled with a cookie in order to track the visitor within the session and possibly across sessions.

### BENEFITS OF TAGGING

You get a number of benefits by utilizing page tagging on your Web site for data collection. If your page tag is designed correctly, you achieve greater accuracy in your clickstream metrics by eliminating some of the effects of proxies and caching such as reduced Web content requests and assimilation of visits. Many network and ISP administrators implement some form of caching as well as use proxies in order to improve the efficiency of the network and have better control and security of the network. These caches deliver previously requested documents without ever notifying your server of the request. A well-designed page tag uses a technique to thwart the cache effect. For instance, a page tag can dynamically alter itself in order to appear to be a new request. Returning the request to your server, the page tag helps you to obtain more accurate counts of requests when proxies and caching are an issue.

If you use page tagging with a session cookie and, potentially, a permanent cookie, you can more easily identify a visitor's session and return visit. You are also able to log metrics like the visitor's local time, screen resolution, and color depth, page title, time on the page, Java capability, and JavaScript capability.

### PROBLEMS WITH TAGGING

Page tagging has drawbacks that you should be aware of:

- Page tagging does not fully capture errors because the page must actually load, and the script must execute in order to capture data.
- Page tagging is heavily dependent on cookies in order to identify the visitor and track the visitor across pages. As more people refuse or delete cookies, this becomes problematic. If only session cookies are used, it is difficult or impossible to identify repeat visitors. Most authorities recommend not relying on third-party cookies,

due to the higher likelihood of being blocked and privacy issues. Even with first-party cookies, you still want to examine whether the tag solution relies on third-party image files or makes third-party HTTP requests, especially if you have strong privacy regulations. If you have multiple domains, using cookies to track users is difficult because the cookie can be read by only the originating domain.

- Page tagging adds overhead to the page-loading times.
- Page tagging solutions require you to wrap non-HTML content like downloads, PDFs, media files, etc. This adds extra cost to the site maintenance and more reliance on the access to the physical content.
- Page tagging essentially fails to deal with browsers that don't accept SCRIPT such as on mobile devices or where security is important.

#### SAS WEB ANALYTICS 5.2 PAGE TAG

New in SAS Web Analytics 5.2 are page-tag templates and documentation on how to include tags in your pages. The page tag is a JavaScript template that sets a first-party cookie on your Web site; the cookie contains the following information:

- Time (also used to foil caching)
- Name of cookie
- Time zone
- Screen dimensions
- Title of page being displayed
- URL
- Referrer
- Whether Java is enabled or not.

The template inserts a request back to your Web server for a 1x1 GIF file. This GIF file is invisible on your page; however, it will add a record in your Web log with the captured information. The request will look something like:

```
GET
/MyWebSite/images/swatag.gif?rndm=1133707330885&swatag=8839120067.8677&tzo=5&whd=128
0x1024@16&ttl=taginternal&url=http%3A//www.mycompany.com/MyWebSite/some/path/catalog
.html&ref=http%3A//www.sas.com/&java=true HTTP/1.1
```

The ETL process can decompose this request as follows:

- rndm=1133707330885 (time)
- swatag=8839120067.8677 (cookie name)
- tzo=5 (time zone)
- whd=1280x1024@16 (screen dimensions and color depth)
- ttl=taginternal (page title)
- url=http%3A//www.mycompany.com/MyWebSite/some/path/catalog.html (URL for visit)
- ref=http%3A//www.sas.com/ (referrer for visit)
- java=true (Java is enabled).

An important feature of the SAS Web Analytics page-tagging approach is that it is a hybrid solution that leverages both page tags and Web server logs. By utilizing both the Web server logs and additional information provided by the page tags, many of the problems mentioned earlier are either eliminated or mitigated.

Because our solution is based on the ETL processing of the Web server log (or logs), we also include directions and guidelines for the ETL administrator. Usually, you process the Web log(s) in the following manner:

1. Convert the logs into a standardized Extended Log Format (ELF). Many application servers have standardized this format.
2. Run a tag conversion script to transform all of the tag data into a standard record format. This step also standardizes the sessioning of the records so that we are able to provide accurate session information.
3. Process the logs with SAS Web Analytics 5.2.

## IMPLEMENTATION METHODS

You can use a number of different methods for adding the page tag to your Web pages. For example:

- You can manually add the SAS Web Analytics page tag to each Web page file.
- You can add the SAS Web Analytics page tag to a template file that each Web page has.
- You insert the SAS Web Analytics page tag into a template file that generates the pages on the server. After the template is modified, the developers then republish the site with the new tag.
- You can configure your Web application server to inject the SAS Web Analytics page tag into each file that it delivers. Most Web applications servers and technologies have ways to perform this action.

Then you configure the Web server to (1) log the tag hits that are standard GIF image requests and (2) to capture the entire GET request. That is, you want the entire query string. Most servers do this automatically. The only additional task you should do is to capture and log cookies. For most servers, this is a simple configuration change with little or no impact on the rest of the Web server.

## SEGMENTATION ANALYSIS

Segmentation analysis in SAS Web Analytics is a visitor categorization or classification method using decision tree analytical techniques. The analysis defines a set of homogenous groups or characteristics (segments) for visitors that have performed a specific action. Segmentation analysis uses the ARBORETUM procedure, a data mining procedure, to produce the analytic segments.

You can use a segmentation report for decision support. By default, SAS Web Analytics 5.2 includes segmentation reports for repeat visitors based on totals and averages. Each segmentation report creates a set of rules that help predict which visitors will return to the Web site. You can then target your Web site to those visitors, or update your Web site to attract less-frequent visitors.

Each segmentation report consists of two tables:

- The first table is the *Segment Definition* table; it indicates whether an analytically-defined segment predicts repeat visits.
- The bottom table is the *Segment Assignment* table; it applies segmentation rules to all of the available data. The Segment Assignment table displays visitor assignments to each segment within the following three time periods: the total time period, after the explanatory time period, and during the current day.

A key benefit of segmentation is the ability to personalize the Web experience in real time by using the segment rules that are provided and the corresponding probabilities. Again, using the default segmentation reports, a segment definition with a large expected response rate (for example, > 0.7) indicates a behavior that can be captured in real time to identify visitors that have a higher likelihood of returning to the site given current and past behaviors. This likelihood might indicate the need for a banner, pop-up, or real-time enticement to motivate people to return, thus producing high-value results and return on investment (ROI).

It is important to note that segmentation analysis requires a persistent visitor ID because it is tracking visitor action over time. While custom versions of this analysis are possible, and indeed desirable, they are not recommended under either of the following conditions:

- If the visitor ID for a Web site is user agent and IP address.
- If a non-persistent cookie is used as the visitor ID.

The analytic segmentation reports included with SAS Web Analytics are examples; study these reports to better understand how to create segmentation analyses that are based on your business needs.

## CUSTOM SEGMENTATION

It is not a trivial task to create a custom segmentation. You must possess some degree of analytical knowledge, an understanding of your data, and the ability to create the data preparation code necessary to transform your data into the form necessary for segmentation analysis.

However, the benefits of custom segmentation are quite remarkable. Consider the following scenario.

You create or select a response variable that corresponds to the purchases on your Web site. You also identify a number of predictive variables including:

- The last time a visitor made purchase (recency)
- The number of times a visitor purchased something in the past 90 days (frequency)
- The average revenue per visit (monetary).

Now, you have an analytical recency-frequency-monetary (RFM) segmentation. The benefits of this segmentation over standard RFM segmentations are that the segments are derived analytically with actual predictive value. In this model, the RFM splits are determined by the analytical decision tree rather than a simple percentage or ad-hoc attribution.

## REPORT AND SUMMARY ADMINISTRATOR

SAS Web Analytics 5.2 provides you with a Web interface for the administration and customization of your ETL processing and report generation. This interface is not as visible as the Report Viewer, but it is very flexible and makes your report and analytical customization dynamic.

### TRAFFIC REPORTS

The Traffic Report Administrator gives you the ability to quickly and easily add new traffic reports as well as customize existing ones. SAS Web Analytics 5.2 adds a number of features to improve traffic report administration. You can now search for specific reports as well as copy and move whole groups of reports in addition to individual reports. When you change a report, the change is immediately reflected in the Report Viewer once the data is refreshed.

When you create a report in SAS Web Analytics 5.2, it defaults to a *rapid creation mode*. This means that the basics are populated for you. This feature allows you to quickly get a report that you can start customizing. In addition, while you are working on a report, you can select that the report is not publicly available so that it does not appear in the SAS Web Analytics Viewer.

As you refine your report, SAS Web Analytics 5.2 provides a preview option with calendar functionality so that you can effectively model your report. A nice convenience in SAS Web Analytics 5.2 is that reports now have an auto-origin feature. This means that—as you test out your report—you will see date ranges that are available for your particular data, rather than having to hunt for the days for which data exist. This auto-origin feature is also available in the SAS Web Analytics 5.2 Viewer and is report-sensitive.

A new feature is the ability to add links based on external reports. These reports can be anything from a link to another application, to a Web site, or to another reporting vehicle. In particular, you can integrate Output Delivery System (ODS) style reports into SAS Web Analytics. An advantage of this system is that your external report can still utilize the calendar component's date selection.

### EQUATION EDITOR

SAS Web Analytics 5.2 has a more powerful Web-based Equation Editor with an enhanced look and feel for building or modifying your reports. The Equation Editor enables you to add mathematically, logically, and data-oriented expressions in order to create a calculated column or add synthetic data to your report.

At a simple level, you can calculate the percentage or other arithmetic quantity from data in your report. For instance, you might want to calculate the average duration in minutes for a visit. The Site Metrics report uses this calculation:

$$({duration} / {session\_count}) / 60$$

This might seem somewhat cryptic. If you observe that the report is based on a data set with variables DURATION and SESSION\_COUNT, it is definitely clearer; however, you might still say to yourself that it must be tedious to enter the expression.

The SAS Web Analytics 5.2 Equation Editor (Figure 11) makes all this much easier.

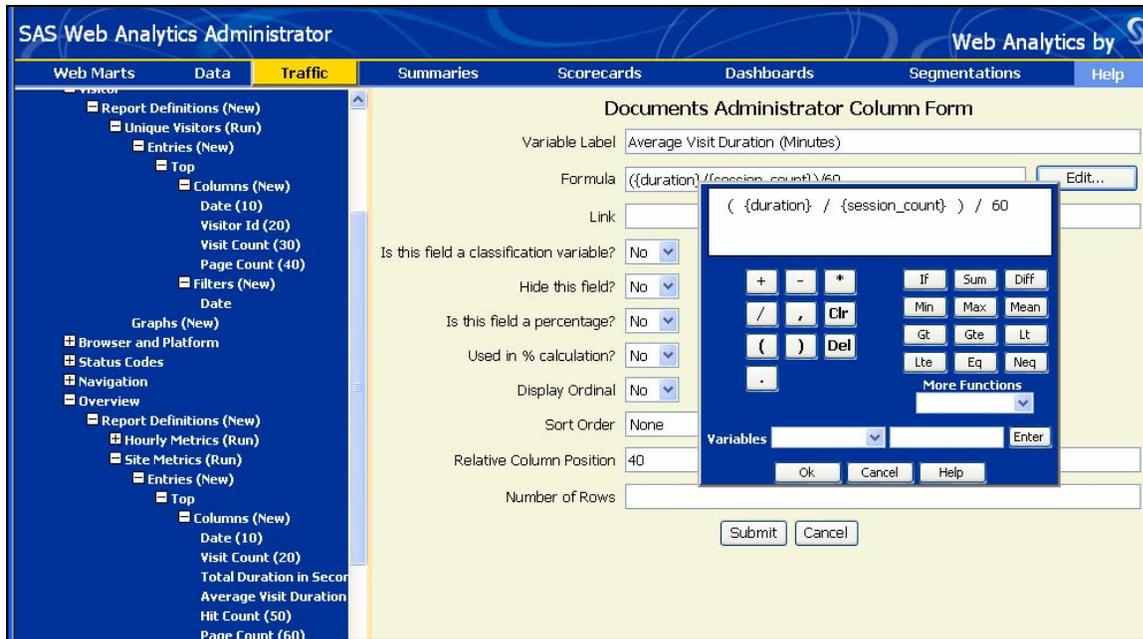


Figure 11. Equation Editor

Much more complicated expressions can be built with the Equation Editor. In particular, first assume that you've augmented the `requested_file` variable with request parameters using a transformation within the ETL. Then you can use sophisticated pattern matching and string operations in order to tokenize the `requested_file` of a Web request and extract the `category_name`. Imagine a store, with different catalogs each containing a set of categories. Furthermore, imagine that a Web request looks something like:

```
/store/product.aspx?catalog_name=books_and_audio&category_name=supplier_relationship_management.
```

If you want to classify your data by the `catalog_name`, define a formula that tokenizes the `requested_file` using the delimiter “=”. Then, select the second element that you want to tokenize using the delimiter “&” and select the first element:

```
token ( token ( {requested_file} , '=' , 1 ) , '&' , 0 ).
```

Remember all this is done through the graphical Equation Editor.

Next, in this report, you might want to further divide the catalog by the category. You can do this by adding another column, with a formula that considers the `requested_file` a property string of name=value properties where the properties are delimited by “&”. We want the `category_name` property, so this formula is:

```
property({requested_file}, '&', '=', 'category_name').
```

An example of this report is shown in Figure 12.

Catalog <i>f(x)</i> ▲ ▼	1-5 Category <i>f(x)</i> ▲ ▼	Session Count ▲ ▼	Hit Count ▲ ▼	Page Count ▲ ▼	Client Error Count ▲ ▼
books_and_audio (1 to 5 of 51)	apqc_publications	8	8	7	0
	best_sellers	10	19	18	0
	harvard_business_review_book_series	4	4	2	0
	leadership	8	8	7	0
	activity-based_costing/management_(abc/m)	7	7	7	0
university_catalog (1 to 1 of 1)	risk_management	1	1	1	0
video (1 to 4 of 4)	corporate_culture	5	5	5	0
	creativity	1	1	1	0
	leadership	3	3	3	0
	stanford_executive_briefings_video_series	2	8	8	0

Figure 12. Example Store Catalog by Category Report

### DECISION SUPPORT

SAS Web Analytics 5.2 has added similar functionality to the decision-support administrators. *Decision support* includes scorecards, dashboards, and segmentation. You can now preview the dashboards and scorecards that you design. Each report must actually execute a decision-support stored process that operates on the defined data sources.

SAS Web Analytics 5.2 has modularized all the decision support processes. This allows you to run, at will, individual or all of decision support reports defined for each group.

### SUMMARIES

The Summaries Administrator allows you to create and customize the summary metadata that will build the defined summary table. At the most general level, this is controlling the SUMMARY procedure with various options (Figure 13).

SAS Web Analytics 5.2 provides new features in the Web-based administrator for working with the summary definitions. You can now copy a summary definition which aids in the creation of multiple similar summaries. You also have the option of selecting whether or not the summary should be run at ETL time. This is useful when you are working on a summary definition and don't want to jeopardize the batch ETL process with a potentially unstable or unproven summary.

Also helping you while you are experimenting with a new summary is the summary test feature in SAS Web Analytics 5.2 as previously discussed.

Summary: Visitor		Modify   Copy   Delete
Label	Visitor	
Source Dataset	session	
Results Dataset	visitor	
Class Variables	visitor_id, date	
Analysis Variables	beg_datetime, end_datetime, page_count, session_count, duration, page_view_le1_ind, page_view_2to4_ind, page_view_ge5_ind	
Summary Levels	day, week, month, qtr, year	
Summary Options	missing, chartype	
Output Data Set Options for Initial Summarization	keep=visitor_id date min_beg max_end page_count session_count duration page_view_le1_ind page_view_2to4_ind page_view_ge5_ind _type_ drop=beg_datetime end_datetime rename=(min_beg=beg_datetime max_end=end_datetime)	
Output Data Set Where Statement		
Output Data Set Options for Re-Summarization	keep=visitor_id date min_beg max_end page_count session_count duration page_view_le1_ind page_view_2to4_ind page_view_ge5_ind _type_ drop=beg_datetime end_datetime rename=(min_beg=beg_datetime max_end=end_datetime)	
Output Options	sum= min(beg_datetime)=min_beg max(end_datetime)=max_end	
ID Variables		
ID Variable Summary Levels		
Input Where Statement		
Run Summary in ETL?	Yes	

Figure 13. Visitor Summary Definition

## ACCESSING THE SAS WEB ANALYTICS CALENDAR

You can use the SAS Web Analytics 5.2 Administrator Web application to create an external report within SAS Web Analytics. For instance, this external report might be a Web page containing a form that invokes a SAS stored process. The external Web page, residing within the SAS Web Analytics report frame, can access the calendar by using JavaScript. You could write JavaScript code similar to the following:

```

var selection = "range";
function
setmodel(model){
    var    calendar = window.top.getCalendar("Traffic");
    if (calendar == null) {
        alert("Cannot find the traffic calendar");
        return;
    }
    selection = model;
    calendar.setSelectionModel(selection);
}
function
runit() {
    var    calendar = window.top.getCalendar("Traffic");
    if (calendar == null) {
        alert("Cannot find the traffic calendar");
        return;
    }
}

```

```

    var    dates = calendar.getSelectedDateString("-");
    var    parms = getParms(window.top.location.search);
    parms["document"] = "SUGI.Status Codes with non contiguous dates";
    parms["datelist"] = dates;
    parms["usecache"] = "false";
    var    href = buildHref(parms,window.top);
    window.top.location.href = href;
    return(false);
}

```

You call the **setmodel** method in order to set the selection model for the calendar—either a contiguous date range or noncontiguous date selection. The **getCalendar** method finds the calendar. You must use the **window.top** methods because you are in a sub-frame.

You call the **runit** method when you submit the form and the following actions take place:

1. Find the Traffic Calendar.
2. Retrieve the selected date strings.
3. Get the current set of search parameters from the top window.
4. Set the document parameter to the Status Codes report within the SUGI report group.
5. Set the datelist parameter to the selected dates. This assumes that the SUGI.status codes report uses the “date in” filter in order to make use of the noncontiguous dates.
6. Set usecache to **false** to ensure fresh data for each run.
7. Set the new URL to the top window location.

This example actually uses some subtle features by utilizing a nonpublic SUGI report group that contains a report named “Status Codes with non contiguous dates.” This report is a copy of the standard Status Codes report within SAS Web Analytics, with the modification of the date filter to use the “in” operator for “datelist.”

## CONCLUSION

SAS Web Analytics 5.2 is a flexible and powerful solution for reporting on and deciphering your Web site’s visitor. The clickstream can be augmented with external data sources to provide a fuller and systemic understanding of your customers and visitors. By linking together the various actions and behaviors of your customers, you discover new cross-sell and up-sell opportunities. You discover new ways of retaining your customers and converting your visitors into customers.

SAS Web Analytics 5.2 philosophically is committed to performance and predictive analytics. Going forward, SAS Web Analytics will emphasize these characteristics even more. This will be even more important as Web data grows to the petabyte range and the competition becomes more intense for high-value customers.

## ACKNOWLEDGMENTS

SAS Web Analytics is a team effort first and foremost. The author thanks Caroline Bahler, Frank Roediger, and Robert Levey. The extended team has been crucial to this project. Said Nechab has been a great advocate and persistent stakeholder. His 1000-meter view of the road ahead has been most valuable. Retha Keyser is the kind of champion for your product that you can usually only dream about. Jerry Hosking has helped make this a better product with her early adoption and frequent feedback. Finally, little of these analytics or the product itself would exist if it were not for Dr. John Brocklebank.

## RECOMMENDED READING

Kimball, Ralph, and Richard Merz. 2000. *The Data Webhouse Toolkit*. New York, NY: Wiley Computer Publishing, John Wiley and Sons, Inc.

Peterson, Eric T. 2004. *Web Analytics Demystified*. Portland, OR: Celilo Group Media and CafePress. 2004.

SAS Institute Inc. 2005. *SAS Web Analytics 5.2: Administrators Guide*. Cary, NC: SAS Institute Inc.

Sweiger, Mark, et al. 2002. *Clickstream Data Warehousing*. New York, NY: Wiley Computer Publishing, John Wiley and Sons, Inc.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Jonathan Polito  
SAS Institute Inc.  
1 SAS Campus Drive  
Cary, NC 27513  
Work Phone: 919-531-2559  
Fax: 919-677-4444  
E-mail: [jonathan.polito@sas.com](mailto:jonathan.polito@sas.com)  
Web: <http://www.sas.com>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.