

It's A Brand New eWorld!

Jay Stevens, Whitehurst Associates, Inc., Atlanta, Georgia

ABSTRACT

"It's a brand new world." "The old rules don't apply."
"You need an e-Intelligence solution designed for the web." The truth is... it is not an entirely new world, most of the old rules do apply, and what you need for e-Intelligence is not a web system but a proven business intelligence / data warehousing solution. The dramatic acceleration of technology changes in the online world has driven the creation of completely new business models, created a completely new customer distribution channel, and has resulted in revolutionary process improvements for many companies. Whether they are a brand new dot com or an old line brick and mortar trying hard to become a hip and trendy click and mortar, they all still need one thing, business intelligence and decision support to understand their organizations, their customers, and their suppliers. This paper reviews the reasons that SAS software is better positioned and more capable of owning the e-Intelligence marketplace than any of its competitors. Whitehurst Associates' Autotrader.com™ (winner of the 2000 SAS® Institute Enterprise Computing Award for Commercial Applications) application will be discussed by way of a case study reviewing the SAS technologies used. In addition, the paper will also look at some of the newer SAS technologies that are finding their ways into large dot coms.

WHAT HAS THE INTERNET DONE FOR BUSINESS?

The Internet has changed everything for business, right? Not exactly. The Internet has been incredibly useful for enabling online interaction with business

NEW BUSINESS MODELS

The Internet has spawned the creation (in some cases) of completely new business models. Companies are now making money with the web in ways that were unimagined only a few years ago. For example, if someone told you 4 years ago that an Internet Service Provider would give away free computers to anyone who would agree to a long-term agreement, you would have questioned it. If someone had told you that...

NEW MARKETING OPPORTUNITIES

The Internet is changing once-local businesses into global distribution centers. *Martha's Handcrafts* on the corner of First and Main streets is no longer restricted to selling its wares only to those who happen to wander by. The Internet has made it possible for Martha to set up shop online and sell to millions around the world. Through the Internet Martha and companies of every size have access to a huge new marketing opportunity.

NEW SALES AND COMMUNICATIONS CHANNELS

For existing companies, the Internet represents a tremendous opportunity for expansion of existing business. There are many formerly offline businesses that have successfully leveraged their strengths onto the web (direct marketing powerhouse Fingerhut, and

publishing giant Ziff Davis). In addition, many companies are taking advantage of and integrating the web and other electronic communications channels into their communications strategies.

THE SPEED OF BUSINESS HAS CHANGED FOREVER

Almost every day we hear about Internet Time or Net Time. The assumption being that, for business, the time available to act and react has diminished by an order of magnitude. To a large degree this is true. The average refresh frequency of data warehouses is dropping. While most companies used to be satisfied with a monthly refresh, many are moving at least some part of their warehouse to weekly and daily refreshes. Many of our dot com clients are insisting that refresh frequencies for analysis of their website traffic move from daily to four times a day and many are looking to get as close to real-time as possible. Speed of implementation and response has increased as well. We made first contact with AutoTrader on a Tuesday. The proposal was in their hands on Friday and we started work the following Monday. Another example of Internet time in action... we had just 4 weeks to architect, develop and implement the baseline system and reports.

MOST IMPORTANTLY: NEW DATA SOURCES

Finally, and perhaps most importantly for those interested in SAS, the web and the Internet have introduced a new data source. For most websites, this data is somewhat unique in that it provides information on the behavior not just of customers but also of potential customers. The most common incarnation of this new data source is the log file(s) generated by web servers (the so-called "web log"). A large website can produce multiple-gigabytes of log files every day. To date, most e-Intelligence activities have revolved around reading, parsing, reporting, and analyzing web logs.

THE OLD RULES STILL APPLY

Even though the Internet has changed a lot of things. For business, many of the old rules still do apply. Even if all of your data now comes from a newer and more voluminous source, you still have to answer the same kinds of questions:

- Who are your customers? Where did they come from? What do they need? What do they want?

- What are they doing with you? How are they buying and using my products and services?
- Why do you care? How does that activity relate to my business model?

E-INTELLIGENCE

At both SUGI and SEUGI this year, Jim Goodnight declared this the year of “e”. The “e” could be found everywhere but was featured most prominently in the magic e-word: e-Intelligence. E-Intelligence now seems to be a new and almost all-consuming central focus for SAS. The word itself is now featured regularly as part of their new logo. So what is e-Intelligence? Let me start by telling you what I believe e-Intelligence is not.

- **E-intelligence is NOT** just learning the startling fact that Internet Explorer now holds 75% of your web audience and Opera only controls 2%.
- **E-Intelligence is NOT** just learning that 85% of your users are AOL subscribers.
- **E-Intelligence is NOT** just knowing that 98% of the visits to your dynamic website were to the base URL:
<http://www.dotcom.com/dynamicpageserver.cgi>

These superficial statistics were interesting to note in the days of the web’s infancy (remember when the most important measure of web success was the “hit” ?) when a website was nothing more than an elaborate marketing brochure. With the introduction of e-Commerce and other large-scale sites, this has all changed.

E-INTELLIGENCE IS B-INTELLIGENCE

E-Intelligence, first and foremost, is actually B-Intelligence. It is Business Intelligence. It must have information value to the business and be useful for decision-making before a business commits to using it.

E-INTELLIGENCE TRANSFORMS CLICKS INTO BUSINESS INFORMATION

The atomic unit of measurement for most websites (eCommerce or otherwise) is “the click”. The click is what drives activity on the web. But these clicks are more than just insignificant transactions that can be summarized into a daily count and discarded. E-Intelligence is the process of extracting decision-useful business information from the click.

E-INTELLIGENCE IS RELATED TO AND DRIVEN BY THE BUSINESS

Whether you are interested in understanding your own internal organization, your customers, or your suppliers, e-Intelligence is about understanding the real business impact of your online interactions with others. Web reporting is nothing if it does not directly relate to your business. Is it useful to generate monthly reports to look at the current state of the “browser war” on your site or the distribution of hits coming from .edu institutions? It might be, but as a rule it is not. We should treat the data pulled from web servers just as skeptically as we would treat

detailed credit card transactions or any other operational business data source.

WHITEHURST FRAMEWORK FOR E-INTELLIGENCE

Our approach to building web traffic warehousing, reporting, and analysis systems for dot coms have been to build reporting and data extraction with the business at their core. Understanding that analysis and reporting should be driven by business needs we have built the following chain of assumptions:

- Clicks are business events
- Business Events drive Business Processes
- Business processes drive the business

CLICKS ARE BUSINESS EVENTS

As mentioned above, the click is traditionally the lowest level of granularity available in web activity data, the atomic form if you will. Most shrink-wrap packages and many higher-end packages have focused on reporting on clicks in isolation. Based on our experiences with websites of all kinds (eCommerce, Content, eMarketplace, etc.), we believe that clicks are more properly treated as business events. In almost every case, it is far more useful to report, analyze, and understand website activity based on business relevant events than by page names. Consider, for example, the utility of each of these mini-reports below:

Figure 1

Page Name	PageViews
/index.html	13,245
/cgi-bin/shopping.cgi?oid=45df32&init=213&user=221	12
/cgi-bin/shopping.cgi?oid=45df32&init=213&user=223	5
/cgi-bin/shopping.cgi?oid=jhj867l&init=213&user=223	18
/cgi-bin/shopping.cgi?oid=45df32&init=213&user=227	8
/cgi-bin/shopping.cgi?oid=45df32&init=213&user=283	3
... etc. etc.	...

Figure 2

Page Name	PageViews
Home Page Welcome	13,245
Shopping Home Page	345
Shopping Recommendations	123

Figure 2 and Figure 1 are technically equivalent but in Figure 2, a business understanding of the technical details of the website has been applied. Figure 1 only

represents data, Figure 2 actually begins to get at some real and useful information. This is the primary difference between web log reporting and e-Intelligence.

BUSINESS EVENTS DRIVE BUSINESS PROCESSES

Business events do not usually live in a vacuum. Each of the business events shown in Figure 2 above have a place in a business process. Once the business events / clicks have been defined, it is useful to map them to business processes. For example, the two shopping pages / events may fit into an overall Shopping process. Once web clicks are mapped in this manner, web pathing analysis takes on a new and more useful business meaning.

BUSINESS PROCESSES DRIVE THE BUSINESS

These business processes are the core of what drives any online business. Even if the business model is just one of trading content for advertising dollars, understanding the ways in which people use the website is crucial to improving the site and its profitability.

WEB LOG REPORTING AT AUTOTRADER.COM

The winner of this year's 9th Annual Enterprise Computing Award for Commercial Applications is Autotrader.com. "AutoTrader.com was recognized for its use of SAS...in developing an interactive management application to collect, process and analyze data and critical site metrics needed to run and grow their business effectively." Whitehurst Associates designed and developed the Management Dashboard application that won this award and is still in use at Autotrader.com today.

AUTOTRADER.COM BACKGROUND

Autotrader.com is the largest used car marketplace in the country. With over 1.5 million used cars and trucks online at any given time, it has the largest online used car database in the world. The online automobile listings are sourced from a network of used car dealers, the listings of Trader Publishing's Auto Trader magazines, and private listings from individuals who visit the site. Listings are free of charge. Its primary business model is an advertising-based one. General advertisers pay for banner and other types of ads, in addition automobile dealers can pay for more targeted ad placement and listing enhancements returned from vehicle searches. Users who visit the site in search of a used car or truck can enter various criteria for the exact vehicle they're looking for (Year, Make, Model, Mileage, etc.). Once a search result is returned, the user can browse through individual listings and get more detailed vehicle information in some cases even retrieving photographs.

ORIGINAL OBJECTIVES OF THE PROJECT

When Autotrader.com contacted Whitehurst, their goals were fairly circumscribed. Their primary objectives were:

- To build an enterprise-strength solution for analyzing and reporting on website traffic and activity.
- The ability to apply customized business logic to the data extracted from the web logs in order to understand the business.
- The ability to integrate the web log data with other internal databases (including the online inventory database).

WHITEHURST APPROACH

In approaching the AutoTrader project, Whitehurst quickly recognized that although the bulk of the data was from a new source (i.e. the web log), the problem was actually an old one. The real problem at AutoTrader was not just "how do we read and parse web logs", it was properly viewed as a business intelligence problem. AutoTrader wanted to understand their web site visitors: their activity and how it related to their business. What was required at AutoTrader was not a new "web age" solution but a data warehousing / business intelligence solution. Obviously these are areas in which SAS has industry leading capabilities.

WHITEHURST SAS SOLUTION

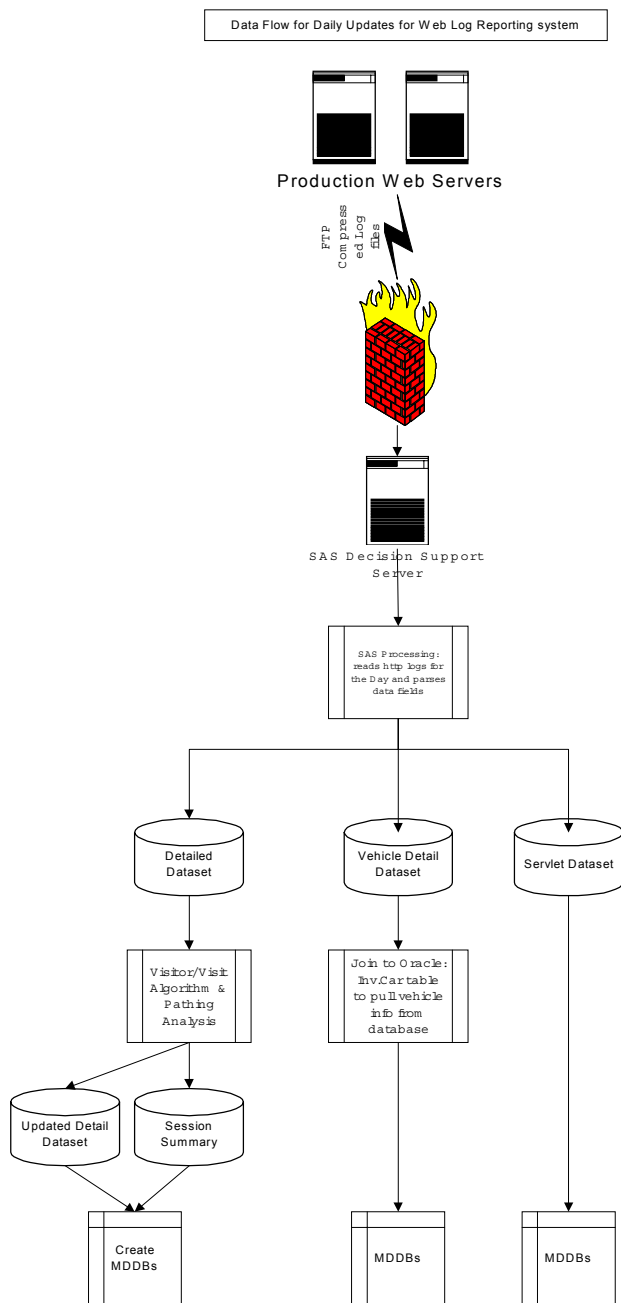
The solution we finally deployed at AutoTrader ran on a Sun Ultra Enterprise™ 4500 with approximately 70GB of RAID disk arrays. From a SAS product mix perspective, our solution used the following:

- Base SAS
- SAS/Graph
- SAS/IntrNet
- SAS/Mddb Server
- SAS/EIS

It is interesting to note here that the bulk of the work for the initial application was done solely by Base SAS and SAS/Graph. The other pieces were only added so that web-based access to the MDDb cubes could be enabled. SAS' cross platform capabilities were also especially useful. Although the system was finally deployed on UNIX, approximately 90% of the system was developed under Windows NT. The migration from NT to UNIX only required 3 days from start to finish.

AUTOTRADER'S DAILY SAS PROCESS

The original daily process of loading and processing web logs at Autotrader.com is shown below.



READ AND PARSE

Every day, beginning around midnight, the web logs are rotated (i.e. the daily file is closed and a new file for the current day is started), compressed, and sent via FTP through the firewall from the Production web servers (originally there were 2 servers, there are now over 10) to the SAS decision support machine. An SCL program handles all the control, flow and execution of SAS code. This SCL program is launched via UNIX cron (job scheduler) around 3 am. The SCL program then begins checking to see if a complete day of log files has been received. If the transmission has been successful, then the SCL program launches SAS code to read and parse the web logs. In this process, variables are created for all of the standard web log fields as well as business specific calculated fields that are parsed from the URL Query

String (the list of name/value pairs that follows a dynamic URL program name). This is the phase where individual clicks are mapped to business event variables and detail datasets are generated.

TRANSFORM AND PROCESS LOG DATA

In the next step of the process, SAS processes the detailed data and generates a session ID that is used to uniquely identify each visit or session that a web visitor makes to the site. A session is generally defined as consecutive website activity for an individual visitor that has no more than a 30 minute window with no activity. So, for example, if a user visits a site, surfs for 1.5 hours, then goes to lunch for 45 minutes, when they return, all of their subsequent activity will be treated as a new session or visit. Because a user's website activity may be spread across multiple physical servers and associated log files, determining session information for most sites must almost always be done as a post processing step. This is one of the most important but time consuming steps in the process. The output of this process is a

PUBLISH REPORTS TO INTRANET

Once the sessioning calculations have been completed, a separate process is launched to generate static HTML reporting. These reports include the main Management Dashboard and its associated network of pages, including daily-generated graphs and links.

BUILD MDDB FOR DYNAMIC OLAP

After the daily dashboard has been updated, the SCL program then launches a series of SAS programs that create and distribute multiple MDDB cubes that are accessed via SAS/IntrNet and the MDDB Multidimensional Report Viewer.

BUSINESS ADVANTAGES

AutoTrader itself has documented several business advantages of the Management Dashboard system with demonstrable financial impacts. These include the ability to surface up to date analysis of traffic volumes to sales representatives out in the field via the MDDB report viewer and their intranet. In addition, the system was used to monitor response to the famous Super Bowl commercial throughout the day. This information "gave them confidence" to continue the offline ad campaign. The management dashboard application also provides insight into the effectiveness of online advertising campaigns, providing detailed information about quality of traffic directed from partners to the AutoTrader site. In addition, the system has been used to track and audit Autotrader.com's large distribution partnerships, most notably with America Online. Overall Autotrader.com has been extremely pleased with their return on investment in SAS.

SAS ADVANTAGES OVER "SHRINK-WRAPPED"

SOLUTIONS

Some of the competition SAS faces in this area is from so-called "shrink-wrapped" products, most notably WebTrends. WebTrends was one of the first movers into the web log analysis space and as such has garnered a great deal of attention. The primary disadvantages of these shrink-wrapped products are:

- **Lack of Scalability:** This is the primary disadvantage that users of these products have related to us. In many cases, it takes over a day to parse, load, and analyze a day's worth of log files.
- **Lack of Flexibility:** Because these products attempt to be all things to all people they lack the flexibility to incorporate customizations for each user such as the ability to parse and define the name/value data available in the URL Query string.

SAS obviously excels in both of these areas and has significant advantages over more narrowly focused web reporting packages. These packages seem to be fine for smaller sites, but tend to fall down when businesses are looking for the flexibility and scalability required for true business intelligence applications.

SAS ADVANTAGES OVER "HOME-GROWN" SOLUTIONS

Another potential competitor for SAS-based e-Intelligence solutions comes from "home-grown" solutions. Because of the exponential growth of the web and the need to rapidly develop reporting applications with whatever tools were available, tools and languages that "cut their teeth" on the web have taken root. These include scripting languages such as PERL and PYTHON as well as lower level languages such as C and C++. The availability of these languages (and full blown databases such as MSOL and MySQL) for no financial investment whatsoever has given rise to a deep reliance on these tools in a number of dot coms. Although these languages have some scalability issues, they do excel in flexibility, providing a high level of control. The primary disadvantage is that they lack a reliable exploitation environment and infrastructure. Every report or change to the infrastructure must be coded in the lower level language. SAS, on the other hand, has a very rich background in deep analytic and information delivery tools that is valuable for any company interested in truly understanding behavior on the web.

THE FUTURE: A WORD ABOUT WEBHOUND

Our solution at AutoTrader was completely custom and used standard SAS products. All of our subsequent e-Intelligence implementations have followed similar infrastructures. At this year's SUGI, SAS announced Web Hound™, their solution for e-Intelligence. For several reasons, we felt that the approach taken by Web Hound 1.0 was not appropriate for a business intelligence application and was more appropriate for a capacity planning/ performance monitoring application. The latest

version of Web Hound soon due from SAS has completely revamped its internal framework. We believe that this is absolutely the right approach and are very excited about using this product with all of our current and future clients.

THE FUTURE: WEB LOGS ARE GOING AWAY

Although the bulk of the e-Intelligence effort has been focused on reading web logs and understanding them, it is our belief that they will soon go away. When one considers that the original purpose of traditional web logs has nothing to do with tracking web site visitors, it almost seems silly that it has persisted as the primary means of recording website activity. With the emergence of Java Servlets and other dynamic server based technologies, we believe that web logs will be forsaken in favor of dedicated custom activity logs, or real-time database recording of web activity details.

CONCLUSION

There is no tool or suite of tools in the marketplace today that can compete with SAS in the world of high-end e-Intelligence. Although it may seem like a brand new world that has no use for a 24-year old piece of mainframe statistical software. The SAS System is still vitally important, useful and critical in today's eWorld.

REFERENCES

SAS Institute Web Site, Enterprise Computing Award press release
(<http://www.sas.com/news/preleases/041000/news4.html>)

ACKNOWLEDGMENTS

The author would like to extend a special thanks to the following:

- Joe Whitehurst, President of Whitehurst Associates
- David Lilly and Matthew Keogler of Autotrader.com
- The SESUG Conference Planning Group (especially Joe Kelley).

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Jay Stevens
Whitehurst Associates, Inc.
341 Eureka Drive
Atlanta, GA 30305
678-358-5734
jay@whitehurst-associates.com
<http://www.whitehurst-associates.com>
<http://www.eintel.net>