

Warehousing Web Analytics

Semphonic in partnership with Netezza



Contents

Table of Contents

- Executive Summary1**
- Current State of the Art – Web Analytics2**
- Current State of the Art – Data Warehouse Appliances3**
- How Traditional Web Analytics Solutions Fall Short4**
 - Volume and Cardinality4**
 - Sampling5**
 - Session-Level Information6**
 - Visitor-Level Information6**
 - Site Taxonomy6**
- Key Benefits of a Web Analytics Warehouse7**
 - Benefit #1: Superior Actionability7**
 - Benefit #2: Advanced Analysis10**
 - Benefit #3: True Customer Research15**
- Conclusion17**
- About18**

Executive Summary

Innovation in the use of online data via traditional web analytics solutions has plateaued – focusing primarily on reporting basic facts about aggregated site usage. From long experience with Customer Relationship Management (CRM) and Business Intelligence (BI), it is clear that aggregate reporting alone represents only a fraction of the opportunity to create value from the collection and analysis of customer behavior. Integrating site behaviors with additional customer information, understanding visitor mindset via current behavior, and driving customer contact via intelligent analysis of that behavior are the applications that drive the most value to the organization. Traditional Software as a Service (SaaS) web analytics solutions do not provide the tools, flexibility, or level of access to support these applications – leading organizations to reconsider exclusive use of the SaaS model for the collection and use of online behavioral data. Bringing online behavioral data in-house has its own set of challenges. The amount of data generated by high-volume web sites is enormous. This vast amount of information is not easy to aggregate in a meaningful fashion and often lacks sufficient metadata to be handled intelligently; similarly, nor are the analytic techniques for understanding and using online data well understood or completely mature.

Concurrently, significant advances in database technology—especially as relates to large data volumes—have made it possible to analyze large data sets without adding complexity and with a lower total cost of ownership than traditional database platforms. The advent of data warehouse appliances such as the Netezza system have made in-house management of analytical warehouses cost effective, easy to maintain, and more flexible than hosted solutions.

This combination of massive data volumes and lack of maturity in handling online data has a significant impact on the type of infrastructure most appropriate to warehousing web analytics data. Huge query volumes require a powerful Massively Parallel Processing (MPP) architecture. But the lack of maturity in handling data means that the most appropriate data model is not well understood. Because of this, any good solution should minimize the amount of data modeling and indexing necessary to achieve good performance. When you don't necessarily know the right questions to ask, the right aggregations to make, or when the "right" question is always changing, it's vitally important to retain the flexibility of a simple, straightforward data model. This places a premium on several key features to state-of-the-art warehousing appliances like Netezza's. There is, after all, no point in warehousing data if getting at the information you need is too slow or too expensive to be practical.

This white paper consists of three major sections: a brief description of the current state-of-the-art in web analytics reporting and analysis, a more detailed exploration of the fundamental data problems in web analytics that cause the shortcomings in the "hosted" approach to web analytics, and finally a section enumerating the key analytic benefits that are enabled by warehousing analytics data. Along the way, the paper will highlight why Netezza's Data Warehouse (DW) Appliances have significant advantages in the web analytics domain.

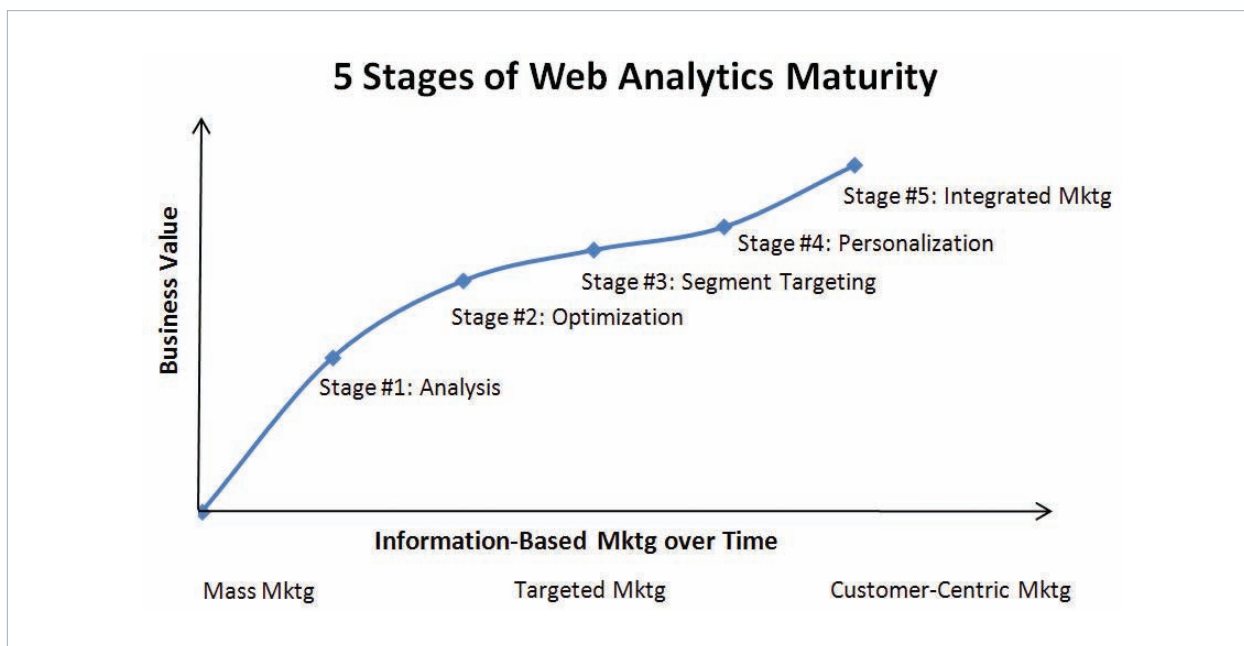
Current State of the Art – Web Analytics

Online measurement has focused on creating a picture of how the web site is being used. Today's current generation of web analytics tools provide powerful insight into how much content is being viewed (page views, visits, uniques), when content is being visited (day, time-of-day), and the relationship between individual pieces of content (pathing, success).

Key Point: Once organizations master the basics of web analytics, they begin to ask questions that are difficult or impossible to answer with traditional web analytic tools.

Until recently, most organizations were struggling simply to make this picture consistent, accurate and reliable. To achieve those goals, organizations had to invest in their web measurement infrastructure, improve the quality of their tools and implementations, and create organizational knowledge about how to interpret basic web analytics numbers.

State of the art organizations in web analytics have now achieved a level of comfort with aggregated reporting on web usage. They are confident that all of their content is being measured appropriately. There is widespread consumption of basic analytic statistics and comfort with concepts like visits and unique visitors that might once have been sufficient. They have at least some dedicated analysts or power users who are comfortable researching specific usage questions using the web analytics tool.



This is indeed progress.

But the state of the art has changed - organizations have begun to discover that there are a great many questions that can't be answered in aggregate using their current tools and siloed data. Often, these aren't esoteric questions but seemingly simple business questions for which having a good answer is imperative. Questions like "How did a visitor's behavior change after they registered on my site?" or "How do my best customers use the web and is it different from the way most of my customers use the web?" or "Which of my products is most interesting to a customer I'm about to send an email to?" have proved difficult or flatly impossible to answer using only a traditional web analytics tool.

Of course, organizations have no desire to give up their clean data, no desire to lose the aggregate reporting they have come to rely on, and no desire to lose the analytic capabilities that are available in current systems. So more and more organizations who have reached the point on the web analytics maturity curve where comfort with basic measurement has been achieved and the problems with analysis have begun to emerge are taking the step of moving the data from the traditional web analytics solution to their own platforms.

Current State of the Art – Data Warehouse Appliances

The capabilities of state-of-the-art database platforms have evolved in parallel with the maturity of web analytics. In-house database platforms have traditionally entailed expense, significant complexity, and the requirement to staff technical administrators to maintain the environment. Even today, the majority of relational databases are based on technology built for transactional applications, not the analytic applications needed for effective web analytics.

All this has changed with the advent of data warehouse appliances, a technology created and brought to maturity by Netezza. The same advantages of web-based software—simplicity, ease of management, and low total cost—can now be found in high-performance data warehouse appliances. Unlike traditional database platforms, Netezza data warehouse appliances are built from the ground up for complex analytics against large volumes of data, large-scale joins, and maintenance-free operation.

Netezza is able to achieve extraordinary results by building a system designed to process data at the speed it can be streamed off the disk. When combined with the power of massive parallelism—100 or more ways in typical installations—this power to operate at the streaming speed of the data is a core advantage of the Netezza architecture.

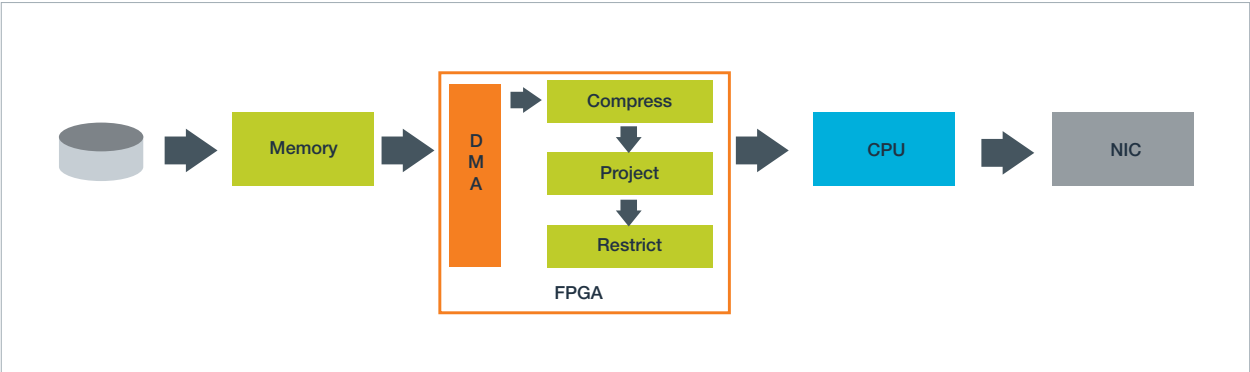


Illustration of a Netezza Processing Unit, one of a hundred or more in a single appliance.

Equally as important is Netezza’s use of the Field Programmable Gate Array (FPGA). This specialized chip is able to handle millions of logic calculations faster than data can be read from disk. The primary purpose of the FPGA in Netezza’s architecture is to remove any rows and columns not needed by a query before it gets to the CPU, and without requiring indexes or partitions—a crucial advantage in web analytics where large volumes of data are routine and queries may not be predictable. By significantly reducing the volume of data before it gets to the CPU, Netezza is able to ensure that only the data required in the final answer is being processed. This unique advantage is the key to manipulating large data sets without introducing the complexity of managing traditional performance mechanisms such as caches, indexes, and partitions.

Traditional databases have proven unable to scale to the volumes required to manage granular web traffic data. As we will discuss, the removal of this limitation presents significant opportunities to modern organizations and represents the next evolution in web analytics.

How Traditional Web Analytics Solutions Fall Short

Some of the classic problems associated with web analytics data have to do with the “shape” of the data itself. At the most granular level, web behavior information is very detailed. And any query that requires detail is subject to performance problems in a web-hosted application or traditional database management system.

Key Point: Most traditional web analytics tools simply won't let you answer any questions about a specific visitor and are extremely limited in their ability to segment and track groups of visitors. These limitations are driven by 1) very large amounts of data and 2) high cardinality in key variables.

Most traditional web analytics tools have focused on a “page-based” aggregation of the data. This means they report nearly everything in terms of the pages viewed (or other site actions like products purchased or registrations made). This works very well for understanding overall traffic and content interest on a web site. But it breaks down whenever you want to understand things from a visitor perspective. Most web analytics tools simply won't let you answer any questions about a specific visitor and are extremely limited in their ability to segment and track groups of visitors.

This is a fundamental limitation when trying to use web analytics data to drive customer contact programs or to execute most types of advanced statistical or time-based analysis.

Netezza Advantage

It isn't easy to solve the twin problems of volume and cardinality. Part of the reason traditional web analytic vendors “cube” data is because traditional data warehousing platforms don't deliver adequate performance when faced with this twin problem. However, with Netezza's use of ZoneMaps™—automatically maintained information about where data lies on disk— high cardinality often results in faster query speeds and better parallelism. And high-volume tables can often be queried in the same or less time as much lower-volume tables.

Volume and Cardinality

It's an unfortunate fact that web analytics solutions intended for big enterprises often have to deal with almost unprecedented amounts of detail data. Big web sites generate billions of rows of data in a month. It's a lot. Even the amount of data for a single day can be enormous. Because the raw data set CAN be so large, tools often can't take the approach of using the low-level data to answer every question. Instead, many web analytics solutions rely on a technique developed well before web analytics to handle very large data sets in other fields. This approach - OLAP - involves the creation of data “cubes” that pre-summarize specific relationships in the data.

But a cube can't capture every possible relationship in the data – if it tries to do that, it gets to be bigger than the original raw data. So the cube builder (in this case your web analytics vendor) has to make some tough decisions about what data to include. In some cases, these decisions are driven by basic analysis of what the vendors think is likely to be important: screen size by referring site – no; conversion by campaign – yes. But there are other factors to consider and the biggest is something called cardinality.

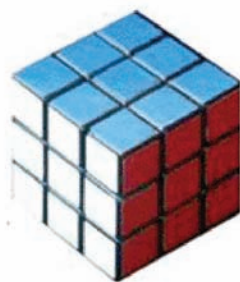
Cardinality is a measure of how many different values a variable can have. OLAP has always worked great for basic customer analysis because most of the variables had very low cardinality (gender – two values; age – no more than about hundred values and often reduced into four or five categories; income – usually reduced to three or four categories). The lower the cardinality, the more data compression you get with a cube and the more variables you can cross-tabulate and make available in the reporting. The higher the cardinality, the less performance benefit a cube will provide unless lots of variable cross-tabulations are eliminated.

Unfortunately web analytics is filled with important variables of very high cardinality. These include page names, content sections, search terms (internal and external), referring sites, paths and visitors. Cubes rely on the effective compression of thousands or millions of records into just a few values, but most of the things you actually care about on the web have thousands, tens of thousands, hundreds of thousands or even millions of different values. It makes cube design both limiting and frustrating and it's why having the tremendous parallel processing power of a data warehouse appliance makes sense.

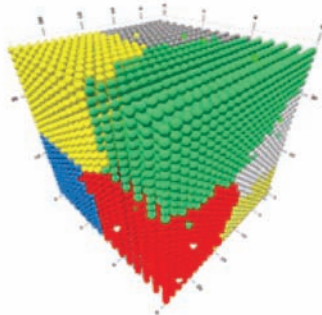
Netezza Advantage

Cubing data can certainly solve certain classes of performance problems, but it essentially removes any possibility of flexible analysis. You can only analyze what's pre-aggregated in the cube. A Netezza data warehouse appliance restores that flexibility by enabling the analyst to ask any question of the data – without limitations – and receive answers with sub-second response time. If you need a team of DBAs to optimize a given access path to the data, then the analyst may have little more functionality in a warehouse than with a cube. Netezza's fundamental technology breakthrough delivers extremely scalable performance without needing ANY database tuning or path optimization...and without adding a staff of administrators.

High cardinality also introduces data latency when depending on cubes or aggregations in a traditional database or web-based solution. The larger the cube or aggregation needs to be, the longer it generally takes to build those structures. Not only does this limit the scalability of a traditional solution as data volumes grow, it also limits the system's ability to support a compelling user experience involving ad hoc, interactive data exploration. Because Netezza's unique architecture does not rely on aggregations or cubes to query the data quickly, this latency bottleneck is eliminated.



Age By Income By Gender



Product By Income By Age



Page Name By Referring Keyword

Sampling

Traditional web analytics vendors provide sampled data for one simple reason – performance. It's much easier to deliver fast answers against sampled data than it is against comprehensive data – especially if you're talking about a large web site with tens or hundreds of millions of requests monthly.

Although sampling is a good alternative for some kinds of advanced statistical analysis, it creates its own set of problems and issues. Sampling can introduce inaccuracies. Many marketing managers distrust sampled data and simply won't use it when it comes time to make decisions. For analytics reporting, sampling is almost always organizationally unacceptable. And, of course, sampled analysis doesn't help when it comes time to integrate your web analytics data with actionable systems for outbound emails or dynamic messaging.

This last point is probably the most important. Many of the biggest returns from web analytics require integration with CRM, personalization, or contact systems. Sampled data can provide understanding, but it can't drive action. When you need to send an email to "Joe Customer," you can't use sampled analysis about visitors because "Joe" probably wasn't in your sample.

Warehousing web analytics data in an appliance solves this problem. Not only does it provide a seamless tie between your customer information and associated online behavior, it also provides easy access to the data. That means you can use online behavioral data to drive email marketing, online messaging, call center conversations, and field sales outreach. Data volumes become manageable because of Netezza's use of the FPGA to eliminate the unneeded records and columns as the data streams off disk. This limits the data that actually gets to a CPU to only the data set required to answer your question. This is so effective that in most cases the size of the underlying table doesn't have a direct impact on the speed of the query.

Session-Level Information

Using a traditional web analytics tool, it may be difficult to isolate user sessions that meet a complex set of filtering requirements (as in, “sessions where X occurred and then Y occurred but Z did not occur, etc”). Summarized data only takes you so far, and there will certainly be occasions when analysis needs surpass the boundaries of these summarization choices. Storing full detail about user sessions, though, will allow you to construct any filter on session attributes without having to be concerned with prior summarization of the data. In a traditional database, adding more filters to your query usually makes the query much slower. In a Netezza data warehouse appliance, adding filters actually makes the query faster. Since the highest-value queries in web analytics tend to also be the most complex, this is a significant advantage.

Netezza Advantage

With most database technologies, adding additional filters to a query make the query slower because the filter needs to be processed in the CPU, stealing cycles away from other analytical computations. With Netezza, the filtering is done in the FPGA—which handles data streaming off of the disk faster than the disk can provide the data. This not only makes processing these filters effectively “free,” it frees up the CPU for other tasks and significantly reduces the amount of data the CPU needs to operate on—a huge advantage for these types of unpredictable, high-value questions, especially as data volumes grow.

Visitor-Level Information

As with sessions, above, it may be difficult to filter visitors who meet a complex set of requirements (as in, “people with characteristic X and Y but not Z, etc”). Again this is due to the use of summarized data. Getting true (de-duped) unique visitor counts across custom date ranges or complex filters is quite a common problem with traditional web analytics tools. And again, the problem can be avoided altogether by querying detailed data in a data warehouse based on an MPP architecture.

Flexibility of visitor selection and the ability to scan many, many millions of rows to identify a small set of visitors rapidly are fundamentally important capabilities for really robust web analytics. The vast majority of questions that marketers want to answer in the real world are about visitors – not about content.

Site Taxonomy

It is important to be able to query web activity data by content, and to drill down using breadcrumbs that mirror the structure of the site itself. Traditional web analytics tools pose a problem, though, because they often require the analyst to decide in advance - at implementation time - what the site taxonomy should be in order to anticipate future reporting and analysis needs. This approach falls short when either the site taxonomy changes, or when analysts wish to conduct a query that is not within the repertoire of reports that were developed in anticipation of needs. There’s no guarantee that you can go back and change report definitions later on, so if you get it wrong the first time then you might be stuck with it. Warehousing avoids this problem by storing full detail data. This allows summarizations on the fly without relying on the data having been summarized in a particular way at implementation time.

Key Point: Even where data summarization is appropriate, being stuck with an inflexible summarization will limit analysis efforts. On-the-fly summarization lets analysts adapt to rapid changes typical of most web sites.

Key Benefits of a Web Analytics Warehouse

In the last section, the basic technical challenges that often drive limitations in web analytics tools and practice were enumerated. But it may not yet be obvious what real-world benefits solving these problems with a solution like Netezza's data warehouse appliance can actually bring. In this section, we'll talk about the real-world applications that flow from an effective approach to warehousing web analytics data.

There are many different benefits to warehousing your web analytics data because there are so many possible uses for online behavioral data. In general, however, three basic categories of benefits will drive warehousing efforts:

1. Warehousing provides dramatically improved opportunities for actionability by joining existing customer data with online behavior and making the combination readily available to a wide variety of marketing systems – from CRM systems to email messaging systems to site personalization systems.
2. Warehousing opens up a number of very powerful reporting and analytic approaches that simply can't be done against "cubed" data or inside web analytics tool that provide little or no statistical analysis capability. Indeed, the term advanced analysis may be an exaggeration since many basic, simple and fundamental questions about visitors and site usage can't be answered in traditional tools.
3. Warehousing is often a prerequisite or an enabler for conducting true customer research using the web. As web populations have become more representative and important, the role of the web in facilitating customer and audience research has grown. A warehouse approach to web analytics data can greatly extend these efforts.

Each of these three benefits will be discussed in more detail in the following sections.

Benefit #1: Superior Actionability

While it is important to get to know your visitors by analyzing their behavior, it's even better to be able to act on that knowledge. The data warehouse approach to web analytics will lay the groundwork for your ability to implement marketing campaigns at every stage of the customer lifecycle, integrate with existing contact systems, and interact with anonymous visitors.

Key Point: Use warehoused data to drive marketing campaigns and complex reporting.

Examples: Full touchpoint marketing, early-stage marketing, at-marketing, re-marketing, behavioral targeting, dynamic messaging, distributed reporting.

Full Touchpoint Marketing

The goal of every marketing organization should be to communicate with the right audience at the right time, in the right channel, with the right message. The combination of message, moment and channel is essential to achieving relevance; in a world where consumers control the conversation, relevance is king.

Web analytics data is your picture of what a customer is doing online – and that online behavior is often the single most powerful tool you have for achieving relevance in your messaging.

The use of web analytics to drive when, where and how you message your audience is the single largest return on investment opportunity in online measurement. Achieving this blend of right message/moment/channel cannot be done except by warehousing your web analytics data and integrating it with all of your messaging systems.

By doing this, you can identify at a very early stage new interests that visitors are developing; you can maximize your opportunities for engagement or ecommerce when a visitor is actively reading or shopping; and you can tailor your ongoing, outbound dialog with customers with a continually updated view of their interests and concerns.

Early-Stage Marketing – Tracking New Interests

Tracking web behavior over time by visitor will provide powerful clues about newly emerging interests and needs. For a media site, knowing that a visitor is becoming significantly more interested in an issue or type of content can make your link suggestions more pointed, help you drive more effectively to registration (by allowing you to tailor your registration value-story to that interest), and help you drive more relevant outbound messaging to already registered visitors. This early-stage interest can be used to drive offers around new products out to the consumer before the actual buying process begins, giving your message a dramatic advantage over competitors.

At-Marketing – Optimizing a Real-Time Conversation

The time when a visitor is actually reading content or actively shopping is incredibly valuable. It's the one time when you know you have the visitor's attention.

Whether you are suggesting new content to consume or additional products to consider, the fundamental advantages and challenges of bringing all of your information assets to bear on the conversation apply. You need to be able to combine what you know about a visitor's previous interests, what you know about the visitor in general, and what the immediate behavior of the visitor is telling you into a single, coherent picture that lets you suggest the right links or the right products for right now.

This benefit is challenging to realize in a traditional warehouse environment, however. Mainstream databases have been built to be queried or loaded, but not both at the same time. In a traditional data warehouse, bulk loads happen in the middle of the night while queries are primarily run during business hours. While a historical data warehouse can be queried in near real-time by a live website, near real-time loads into that warehouse are required to realize the full potential of optimizing a real-time conversation. Netezza has been built to load and query data concurrently—even in the same table. By eliminating the database “locking” inherent in other platforms, real-time conversations not only become practical, but valuable and interactive as well.

Re-Marketing: Tailoring the Ongoing Dialog

When your website enables visitors to create accounts and log in, you can track them throughout their online journey – even when they aren't logged in. This makes it possible to use online behavior to continually tune your outgoing messaging for maximum relevance. In the digital media sector, achieving this level of integration can be challenging. Getting visitors to register is job one if you want to drive an ongoing dialog that isn't bound to the real-time conversation. It's hard work getting that registration, but once you get it, you often have a much more detailed and up-to-date view of visitor interests and concerns than do your counterparts in other industries. To take real advantage of that quality, currency and breadth of information you MUST be able to use it to time and tune your outbound messaging.

Netezza Advantage

The granular data required for effective visitor-level analytics can be a challenge to manage and query in a traditional database, especially when you need to join many large tables together to get an effective view of a customer's activity.

Netezza's use of ZoneMaps and the FPGA make these high-value, high-volume joins not only possible, but orders of magnitude faster than they would be on any other platform.

Regardless of your industry, finding the right message for an individual visitor is typically a function of your ability to summarize their previous behavior versus their most current viewing sessions to identify “magic moments” of new interests. Getting the right message also requires the ability to segment your audience appropriately so that every offer is surrounded by creative that will resonate with their attitudes and interests. Both functions demand high-powered access to web analytics information stored at the visitor level and aggregated by interests. This only happens in a data warehouse.

What Digital Media Can Learn from Financial Services and Retail

Digital media companies can learn much from their counterparts in other industries that have invested in the customer relationship management game over a period of many years. Financial services marketers, for example, place a huge premium on customer acquisition because they think in terms of the lifetime value of a customer versus the short-term value of a visit.

This is a huge change for digital media, but it's also a huge opportunity for value creation that becomes possible when starting down the path of warehousing your web analytics data. Warehousing web analytics data lets you track at the visitor level – not just the content level. Suddenly, you can see when a specific visitor changes interests, does something new, or does something less.

SEM and PPC Arbitrage

Digital Media websites often use Pay-Per-Click (PPC) advertising to drive incremental traffic. But knowing how much you should spend on different keywords isn't a simple question to answer. Not every website visitor is of equal value, and the immediate impression value gained from a visit may only be half of the overall equation. Ideally, you need to be able to understand the Lifetime Value (LTV) of each new visitor to accurately forecast how much you should spend to acquire them. With the full over-time, visitor-based view of traffic accessible in the warehouse, you can leverage predictive analytics techniques to optimize your PPC portfolio and maximize the return on your SEM budget.

These indications of change are critical in customer relationship marketing (they are often associated with the “magic moments” mentioned earlier in this paper) and the ability to identify these changes is an essential part of what warehousing online data provides.

Often, the best indication that many companies have of a “magic moment” is when a visitor visits a new area of their site.

A pure prospect is much more likely to start a search on Google than on a specific branded web site. But an existing customer/reader is quite likely to start with your web site – not a broad internet search. Treating these existing customers as if they were anonymous prospects is bad business; it's bad for the relationship and it's bad for your bottom line. Unfortunately, while existing web analytics solutions allow you to identify some of these magic moments, they have two failings that make them much less effective than they should be:

1. The vast majority of traditional web analytics solutions will only key data based on a cookie. Even if you pass your customer id (or some proxy) to the solution when a customer logs in, your visitors won't actually be keyed on that id; it's just stored as an additional variable. With cookie deletion and rejection rates, this can make tracking visitor behavior over time and across sessions extremely problematic – you can either miss records or create unnecessary duplicates. Using a warehouse, you can re-key data to bind together pages and visits that most traditional web analytics tools will split apart. This re-keying of data is very processor intensive, however, and very well matched to Netezza's massively parallel architecture where all processing is spread across hundreds of parallel CPUs.
2. You can use the tracking of content over time at the visitor level to quickly target visitors whose interest is NEW. Using a traditional web analytics tool, you can often target anyone who visited an area, but you'll have a much harder time selecting visitors who have not visited that area before or (especially given the cookie problem) messaging to them based on the extent and duration of their interest.

The ability to use multiple keys to bind together visitor records (the web analytics cookie, your customer id, email address, etc.) is a powerful tool for extending the reach and accuracy of your customer profiles. It's something that nearly all sophisticated marketers in traditional customer database applications are already doing and it's equally important for digital media sites entering the warehousing arena to address. Tracking online visitors by your key – not just the web cookie - translates into direct improvements in your ability to market to your customers when they are experiencing the “magic moments” that will drive your early-stage, real-time and ongoing messaging.

Behavioral Targeting and Dynamic Messaging

The more you can reduce the cycle time on analysis and visitor tracking, the more messaging opportunities you open up. Although real-time systems exist for certain classes of behavioral targeting, many dynamic messaging opportunities are better delivered by custom targeting and custom serving within your existing IT framework.

The single most valuable time you have to converse with a visitor is when they are actively browsing your site, talking with you on the phone or otherwise immediately engaged with you. During these times, you have direct access to the visitor's attention. With "push" techniques like email, you just can't guarantee that.

If relevance is a combination of right channel, right message and right moment, the right moment is nearly always as close as you can possibly get to the time when a visitor's interests are revealed. Ideally, that means responding to those interests in the same conversation (verbal or online) – not in the next conversation or three days later via an email.

Reducing the cycle time between when a behavior is revealed and when you can message based on this revelation makes new methods of communication (such as on-site messaging) possible; it also makes every message you do send that much more effective. Unfortunately, most databases require significant time to create aggregates or cubes before the data is ready for consumption. Netezza's "always on" architecture—which allows for concurrent loading and querying of data as well as extreme performance even on granular data—allows you to use the discovery immediately, when it will be the most effective.

This simple fact is even more important in the world of digital media than it is in other industries. An interest in gardening or physical fitness or retirement planning will often exist for years or even a lifetime. But an interest in the Beijing Olympics, Obama vs. Bush, or the protests in Iran will often exist for a few short days or hours.

Distributed Reporting

Traditional web analytics tools are good for many types of static reporting, but they don't meet every need. No web analytics tool available provides the kind of data access that a SQL database does.

Moving data into and out of web analytics systems is cumbersome. Data transformation inside web analytic systems is either nonexistent or very limited. What's more, complex distributed reporting needs are often so unique to a business that they simply aren't supported by the level of automation available in traditional web analytics tools.

To automate and distribute hundreds of customized reports to affiliates, franchisees or content contributors is exceedingly difficult and sometimes impossible with a traditional web analytics tool. And while the top web analytics tools provide reasonable export capabilities to Excel, Excel is not able to deliver the complex customized reporting of products like Business Objects or MicroStrategy.

Storing web activity data in a warehouse makes it possible to layer on very powerful reporting tools or take advantage of SQL's inherent strengths to develop sophisticated and fully automated systems for large-scale report creation and distribution.

Using a data warehouse to support more relevant and immediate marketing, to control dynamic messaging and behavioral targeting on-site, or to meet large-scale reporting automation needs are all high-value applications. These types of benefits are often the lynchpin of data warehousing efforts, but they do not by any means exhaust the benefits to be achieved from those efforts.

Benefit #2: Advanced Analysis

Web analytics has evolved its own standard set of reports and tools. Some of these, like pathing, are particular to the web. But most are relatively common in many forms of traditional data analysis. A great deal of web analytics reporting involves simple frequency counts and basic cross-tabulations (e.g. one variable by another variable as in pages viewed by referring site). At the very high end of the traditional web analytics market, the most sophisticated web analytics tools provide n-dimensional reporting – the ability to look at many variables cross-tabulated with each other (as in pages viewed by referring site by Direct Marketing Area (DMA) or city). N-dimensional analysis becomes much more complex as data volumes grow or the number of dimensions increases, primarily due to the time needed to pre-calculate the analysis in multiple dimensions. The Netezza data warehouse, however, removes the requirement to pre-summarize the results. The impact of the ability to perform this analysis in real time cannot be overstated—it's the difference between having the required information and insight when it is most needed, and having completely missed the opportunity.

N-dimensional reporting is probably the single most common and one of the most powerful features in the world of data analysis. But it is by no means all of the functionality an analyst would like to have.

Time-Based Analysis

N-dimensional analysis is great, for example, at taking a snapshot of existing behavior. But it doesn't provide much help in understanding how individual users or groups of users are trending. But behavioral trends are one of the most important things you can understand about a visitor: is their usage going up, going down or staying constant. For any kind of retention analysis, attrition analysis, or long-term customer analysis, understanding the changes in type and quantity of visitor behavior is essential. Web analytics tools simply don't provide any time based or time-series analysis.

Key point: Transcend typical cross-tab reports and conduct more advanced analysis.

Examples: Time-based analysis, pre/post behavioral analysis, segmentation, scoring, data visualization/exploration, predictive modeling.

Digital Media Example: Usage, Direction and Velocity

For digital media, customer mindshare is the ultimate goal. Whenever your web site (or mobile application) is an ongoing touchpoint that you expect and want visitors to use on a regular basis, then there is a need to measure the direction and velocity of actual usage at the individual level. It's great to know that overall usage on your site is increasing – but these macro trends aren't really actionable. It's much more important to know which visitors to your site are beginning to slacken in their usage (you may be losing mindshare), which are in the early phases of usage growth (possible engagement opportunities to lock in additional behaviors), and which are using the site/service less than they ought to.

Taking the knowledge of usage down to the individual level makes the information profoundly actionable – you can now target content and messaging to the right visitors at the right times.

In a data warehouse, you can aggregate and report on individual usage by time period (and even by content or application type). Tracking overall usage, for instance, you could take a twelve month moving average of visits by month. Subtracting the current month usage from the moving average will give you a positive or negative number that represents the size of change. Dividing this size of change number by the calculated moving average will give you a rate of change so that random variations in heavy users don't swamp smaller and mid-size users with significant variation. Calculating rates of change over multiple periods will give you a velocity curve. If you have enough periods (perhaps measuring at the weekly level for frequently accessed sites or services) you can also calculate the standard deviation of the changes by user. Comparing the current deviation from the standard will give you a measure of how significant this particular visitor's change in behavior actually is. This will keep you from messaging inappropriately to users who just happen to display regular variation in their site usage.

These numbers – size of change, rate of change, and significance of variation – are simple to use, and quite powerful. They make it easy for marketers to target segments whose interest is growing or decreasing in significant fashion. Building this kind of power into data makes it much more likely that marketers will really take advantage of the information – and it's something that can only be done by warehousing your web analytics data. Using a Netezza data warehouse appliance, these calculations can be done in real time instead of being pre-calculated in advance.

Pre/Post Behavioral Analysis

Time-based analysis isn't just about individual user trends and velocity. It can be worth examining the change in behavior of a group of individuals both before and after a certain event - a registration, a purchase, the posting of a comment, etc. - to assess the impact of that particular event on the behavior of the individuals. In other words, how do you know if a certain action affects a user's subsequent behavior? The answer to this question lies in the ability to do pre/post behavioral analysis.

Traditional web analytics tools are very limited in their ability to do this - their cube nature makes analysis of specific visitor groups and time periods extremely difficult.

Within the warehouse, however, you can segment a population and extract all of their behavior before and after an event. You have complete control of the time periods involved and you can use a rolling window (in other words, you can extract a specific period around each visitor's individual registration and compare their before-and-after behavior regardless of when they actually registered). This is quite impossible to do within a cube-based analytics system.

Netezza Advantage

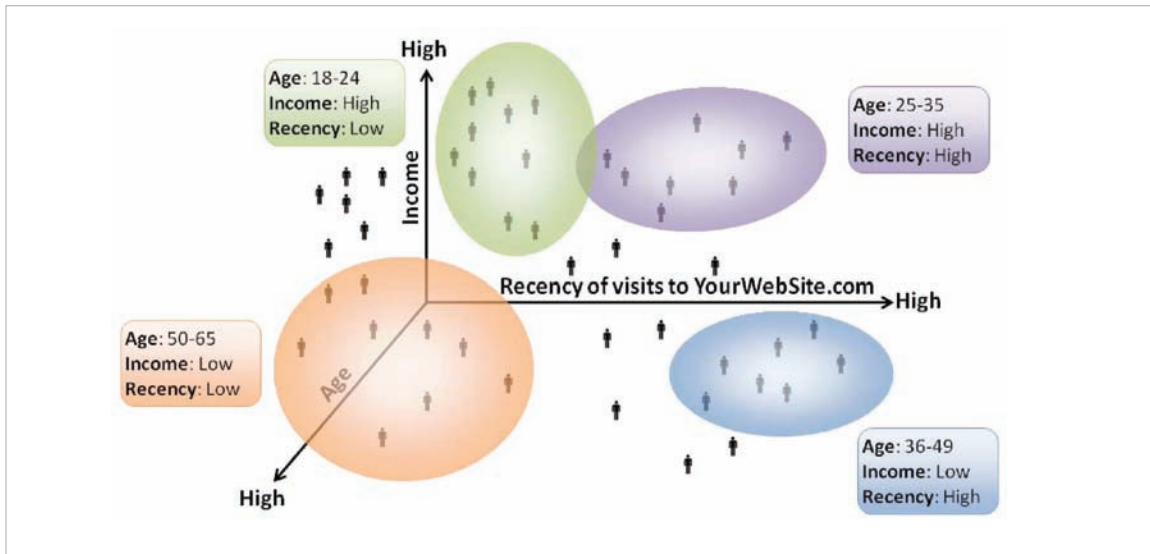
These types of queries can even be very challenging for a traditional data warehouse since they often involve very large and complex joins of visitor data.

This is a case where the ability of the FPGA to discard rows and columns not needed in the solution set as the data is streamed from disk can make a profound difference in performance.

Segmentation

In the broader marketing world, one of the most important techniques analysts use is segmentation: grouping customers into discrete segments based on their demographics, attitudes or behaviors. Segmentation is so powerful because it uses lots and lots of dimensions at the individual level. You may think n-dimensional analysis can do this, but it can't. Not only does n-dimensional analysis operate in aggregate, but it is limited in the real-world to looking at three or four dimensions at most. Why? We poor humans simply can't take in a six-dimensional view of the data; it's just too many cuts for a person to keep track of. Not so a computer. Using statistical analysis tools (like SAS, SPSS, or R), the analyst can use segmentation techniques (cluster analysis, factor analysis or neural networks) to find unexpected, interesting and highly marketable customer segments.

These methods work by combining many dimensions of visitor behavior to map each and every site visitor into a discrete segment. A segment mapping like this describes a visitor across the entire range of their behavior and known characteristics. It can combine site viewing with demographics, geographics, core company relationship information, and even online survey data.



This just can't be done with any of today's traditional web analytics tools – not even the most powerful.

Real segmentation is not simply a matter of selecting a group of visitors who demonstrated some behavior. Despite the confusing terminology adopted by the web analytics world (where cross-tabulations are “correlations” and filtering is “segmentation”), visitor segmentation is about the classification of all the visitors to your site into discrete groups based upon the entire set of their behaviors. This type of segmentation – fundamental to offline marketing analytics – is the single most effective technique for translating the bare-bones numbers of web measurement into the flesh-and-blood understanding that marketers need. Netezza's ability to perform advanced computation on each individual processing node makes segmentation practical on very large data sets.

This type of sophisticated segmentation should be used in nearly every marketing situation to understand the target audience, the success of the initiative by audience type, and the creative messaging (and variations) appropriate to the initiative.

This is computationally a very intense process and it also requires a significant amount of ETL logic. Traditional web data structures don't lend themselves to what is, after all, a visitor-level analysis. So before you can turn an analyst loose on a segmentation, you've got to force the data into a set of visitor aggregations that capture as much of the interesting behaviors as possible.

Netezza Advantage

ETL (Extract-Transform-Load) is the bane of most data warehousing systems. You need to move data from your raw sources into the warehouse and you often need to change it significantly to get it into a structure that is optimal for data access. This takes a whole new set of tools, expertise and time.

While it is possible to use traditional ETL tools with Netezza, the performance impact of hundreds of parallel processors allows organizations to take an ELT approach, loading the data first, then performing the transformations in Netezza. This results in much faster transformation and load times, and eliminates the necessity to invest in dedicated ETL tools and hardware.

Creating a truly online visitor segmentation helps build a powerful connection between web analytics data and marketer decision-making. But there is an enormous additional advantage to doing this type of segmentation in the warehouse.

If you create discrete segment assignments for every visitor, you have a built-in mechanism that is both simple and powerful for customizing the creative of every touchpoint message you generate. This transforms the segmentation from a research exercise to an integral part of your messaging strategy.

An online visitor segmentation code can be used in email marketing to choose the most appropriate creative approach; it can be used in campaign reporting to track the types of audience campaigns source; it can be used by account managers to structure the conversations with key accounts; it is, in short, the key to personalizing messaging creative and not just offers.

The combination of right message and right moment is the cornerstone of achieving marketing relevance in today's consumer driven world. But right message isn't just about offer. A rich audience segmentation is the single most important ingredient in personalizing the tone and creative of every message you make, which ultimately results in dramatically increased response rates.

Scoring

The whole point of getting all this data is to use it. And one of the primary uses of online data is to sharpen your messaging to customers when you send out emails, get calls in the call center, push out a mail piece or build a search campaign – the type of superior actionability described above as the most important benefit to data warehousing.

In the offline world, database marketers have built up a body of techniques that allow them to score visitor interests, predict response rates to these offers for key customer segments, tune messaging strategies, and model how different contact strategies and expenditures will drive different results. These predictive modeling techniques are essential to many advanced database marketing techniques using online data.

A visitor (or visitor segment) scoring system will let you assign values to actions – and those values are additive. With this method, it's possible to produce a score based on the number of product content views, or the total time in a content area, or on a combination of those two values that exceeds a set threshold. Scoring is rarely available in a traditional web analytics toolkit due to the computational power required, but it can be incredibly useful for certain tasks.

What tasks are these? One of the obvious ones is measuring engagement, since it's commonly thought of as a composite measure based on a custom set of actions. CRM integration is another significant application; scoring methods can be used for customer contact programs that are driven by site behavior.

A travel site, for instance, could easily establish threshold values for receiving an email alert on a particular destination. That threshold might include any combination of actual trips to the destination, planned trips to the destination and actual trips to similar destinations. Additionally – and this is a computationally more difficult problem – suppose the travel site sends customers a once-monthly newsletter and wants to target a dynamic offer to the destination they are most interested in. The scoring approach can be used to drive marketing initiatives such as these.

Scoring systems are not available in most web analytic systems or traditional data warehouses with large volumes of data, but they are easily implementable in a Netezza data warehouse appliance. As an additional benefit, scores that you build in the warehouse for targeted messaging often turn out to be useful parts of more generalized reporting and analysis. Netezza's massive parallelism combined with the ability to create custom scoring algorithms in C++ and run them directly on the processing units creates a compelling marketing advantage.

Data Visualization and Exploration

Traditional web analytics tools have made good progress in recent years to provide data visualization tools that meet the needs of the analyst. However, they're still no match for tools that are custom-built for data visualization rather than for web analytics – tools like Tableau or SAS. Warehousing the data makes it possible for an analyst to layer on any best-of-breed third party visualization tool of their choice.

One of the benefits of data visualization is the ability to ask a question, refine it, revise it, and so on until an answer is obtained – this process amounts to data exploration. With full detail data, the exploring analysts face no restrictions on the types of questions that they ask, or the process by which they refine their queries, as long as system performance is not an obstacle.

System performance is not just based on the size of the data set, but also on the complexity, filtering, and join criteria of the queries. Traditional databases can only perform well on large volumes of data when specific questions have been asked—questions that have been optimized for in advance using indexes, caches, or partitions. This requirement to pre-define queries is in conflict with the key value proposition of data visualizations—to explore data in multiple ways to find new patterns. Netezza's use of FPGAs, ZoneMaps,

record restriction, row projection, and massive parallelism instead of traditional indexes and partitions means that any question asked can be answered very quickly. Not only does this mean significantly less administration and maintenance overhead, it makes effective data visualization possible!

Predictive Modeling

As an organization grows into a cycle of data-driven messaging, patterns critical to success begin to emerge. How often a visitor can be messaged, what channels, and what creative strategies are most likely to work will all begin to fall out of the ongoing experience. Marketers in direct response disciplines have learned that these types of patterns can be profitably used to predict which customers/prospects are most likely to respond positively to specific offers/strategies and how to maximize the profitability of existing campaign.

Warehousing data makes this type of predictive modeling practical. First, it provides a platform for the integration of all the necessary components – including your offline data. Second, it gives you direct access to the data at the most detailed level – essential for most predictive analysis. Finally, it opens up the data to the types of tools and programs that are required to build and track these models. Netezza has certified integrations with industry-leading predictive analytics tools including SAS, SPSS, and R. In addition, Netezza's OnStream Analytics capability enables high-performance execution of predictive analysis algorithms running in parallel on hundreds of massively parallel processors.

Benefit #3: True Customer Research

The web has become a significant touchpoint of customers and prospects, and it can be used to better understand your audience: their interests, likes and dislikes, and their affinity for different creative messages. This type of information is quite valuable even when you aren't using it to drive specific messaging opportunities. In short, the web can be a platform for powerful and inexpensive customer and audience research that support both the online and offline channel.

In the early days of the internet, online customer research was hamstrung by the fact that the population you could reach online simply wasn't representative. That isn't true anymore. The web offers a powerful two-way communications resource that you can use to understand your customers, and the learnings you gain there can be powerfully applied to EVERY part of your business including its offline components.

Key point: Use visitor-level detail data to better understand your audience.

Examples: Interest profiling, online panels, complete view of customer attitudes/behavior.

Digital Media Example: Profiling Interests and Mindshare at the Visitor Level

Most media organizations have become skilled at tracking and measuring data about themselves – their inventory, their costs, and their content. Few organizations have developed real skill at attaining a similar understanding of their audience. Yet it is by transforming the information view from content-focused to audience-focused that the greatest opportunities will be unlocked. What digital media outlets sell is not content but audience – access to the right audience at the right time with the right channel. Convincing advertisers that you have that combination of audience, timing and channel is an exercise in online measurement.

How does a digital media company change its information view from content-based to audience-based? It begins by recognizing that your fundamental view of the data cannot be at the content level. Understanding how many people view a page (impressions) is simply not the same and nowhere near as important as understanding what pages a visitor views (audience profiling).

Impressions have become a commodity item. The internet is awash in impressions. The key to selling ad inventory is delivering audience and the investment you make in understanding and describing your audience is the single most important information investment opportunity you have.

In a data warehouse environment, you can track the ongoing usage and interests of every visitor. Typically, this means building up a profile of how much content each visitor viewed and how much of each type of content a visitor viewed. In the data warehouse, this would typically be represented by a single row of information for each visitor / time period containing columns that track how much of each content type a visitor has viewed.

If you are profiling users by the content they view, then it's obviously critical to put real effort into a very robust content taxonomy. The finer-grained your content taxonomy, the better and more descriptive your visitor profiles. For this reason, it's important to systematize, with strong editorial guidelines, your approach to setting and capturing a rich content taxonomy (or moving that taxonomy from the content management system to your warehouse). It's also a good idea to take advantage of nontraditional content hierarchies based on SEO metadata (your editors probably already classify content with SEO keywords) or user-generated taxonomies.

With this seemingly basic profile data comes a tremendous amount of interesting information about a visitor: you know how wide their interests are, you know what topics are most important to them (mindshare), you know how their interests are changing and evolving, and you know how likely they are to stay focused on a "core" interest and how likely they are to get engaged with a current-interest topic. This data can be used to develop a compelling profile of your audience – a profile that can be used to sell audience, not impressions.

Creating Your Own Online Panel

Panels are classic primary research techniques. A panel is group of people who have expressly opted-in to allow you to follow their behavior or survey them on a regular basis. Panels are used by all the large research companies and they form the basis for measurement of everything from TV viewing, to packaged goods brand awareness, to the usage of the internet and its major properties. There's a chance that your organization already uses data from one or more of these types of broad-based consumer panels. But building your own panel has unique advantages and it's something that can only be done in the context of a data warehouse.

Creating your own opt-in online panel is quite inexpensive. It provides you with a platform from which you can measure a whole class of behaviors and attitudes that are nearly impossible to study in pure web analytics. If you want to track the long-term indicators of growth or attrition, a panel is a great way to do it. Because panel members opt-in and self-identify, you can usually put in place mechanisms to track all their behavior – online and offline and without the considerable inaccuracies of any cookie-based tracking scheme. Opt-in panel members step up, not just to letting you track all of their online behavior, but to regularly filling out surveys that can be more extensive, more probing, and more comprehensive than anything you could reasonably attempt with a fly-by instrument.

Because panel members can be tracked regularly, you have the ability to match attitudes to behaviors from an ongoing perspective, not just a snapshot. This can give you means of tracking how external marketing factors are (or are not) moving the needle on the web and influencing online marketing programs like PPC or display.

Your own panel is the best research vehicle for answering questions related to multi-channel interaction, brand impact, competitive pressures, the influence of exogenous factors like the macroeconomic environment, and anything related to the long-term behaviors of customers.

To build your own panel, you'll need an opt-in mechanism, a user tracking mechanism, and a data store that captures and integrates the panel's online behavior, survey results, and – where appropriate - offline behavior. For Digital Media, that means knowing as part of your core analysis whether a panel member is a subscriber to your print media, watches an associated broadcast program, or otherwise accesses your content offline.

You'll want to collect that information for as long as you have the panel – no artificial 1 month or 1 year cut-offs. Because a panel is so much smaller than the total population, you can combine every type of information you have in great depth and study it with the most sophisticated exploration and analysis tools – all without incurring any great cost.

Complete View of Customer Attitudes and Behavior

It's generally recognized that web behavior data is a good indicator of "what" people are doing on your site, but it doesn't necessarily tell you "why" – for that you'll need to use attitudinal data. The most common way to incorporate attitudinal data with web data is to use an online survey, and this integration is fairly easy to accomplish with today's web analytics tools.

But what today's analytic tools don't really allow you to do is analyze the combination of survey data and behavioral data in tandem. In most cases, these tools make you export your analytics data to the survey provider. But the software used by survey providers isn't really designed for behavioral stream data and the analysts at those companies are much more experienced in survey analysis than behavioral data. The result is that very few companies who run online surveys and web analytics tools ever combine the two in an actual analysis.

Warehousing analytics information in Netezza lets you combine these two streams – along with every other important source of individual-level information – and then analyze the information effectively.

Conclusion

The reasons for warehousing analytics data are many and powerful. The need to improve the actionability of measurement and analysis is manifest to most organizations and represents the biggest information opportunity available to most online properties. The opportunity for more advanced analytics and the possibility of better audience research, both to support actionability and to extend the fundamental knowledge of the organization, are often very significant additional benefits to data warehousing.

By warehousing web analytics data, an organization can dramatically enhance its ability to identify key audience segments and target them appropriately. From identifying early-stage product interest and taking command of the sales conversation before competitors can get involved; to creating a highly-personalized on-site experience around shopping, social networking, or lead-generation; to finding and taking advantage of re-marketing opportunities, a warehouse is the key tool for translating online measurement into actionable strategies.

The gains in traditional analytic understanding and audience research are almost as dramatic. Almost the whole of the advanced analytic techniques developed in the offline world to support targeted marketing have been ignored in the online world and are unsupported by common enterprise web analytics tools. From predictive modeling to customer segmentation to time-series analysis and data visualization, warehousing opens up a range of opportunities that simply don't otherwise exist. These techniques are essential for changing the fundamental nature of your online measurement usage from reporting to analysis.

As with most such applications, the justification for warehousing online measurement data is likely to be a combination of hard returns around specific action opportunities, as well as the potential for new uses of the data and new wins created by better analytics and research.

The great challenge, however, is finding an appropriate technology platform. Web analytics is famous – perhaps notorious – for the amount of data its practitioners capture. Compounding the problems always introduced by large amounts of data are the high-levels of cardinality in key web analytics variables and the lack of maturity in the field. Data models for web analytics data simply aren't well understood.

These three factors: quantity of data, high-cardinality in keys, and lack of maturity in data model place a premium on the capabilities delivered by the Netezza data warehouse appliance. Netezza's architecture delivers the core capabilities essential to solving the problem: the ability to handle very large amounts of data; the ability to do high-performance joins on data with very high-cardinality; and the elimination of nearly all of the complex tasks around data modeling. Even better, it does all this while delivering a platform that provides direct access to the data both from SQL and a variety of industry-standard tools essential to delivering better analytics and integration with "action" systems.

For marketing departments that have opted for traditional web analytics precisely to avoid the hassles of traditional data warehousing, the most significant point of all may be the dramatic reduction in classic database administration and IT support necessary with the Netezza data warehouse appliance.

The benefits of warehousing web analytics data are probably obvious to any organization with experience in online measurement. However, past experiences with the challenges and costs of traditional data warehousing have made many organizations reluctant to embrace this direction. For high-volume web properties with significant opportunities in actionability, analysis or research, the Netezza approach offers the potential to have the best of both worlds.

In today's world, there is no middle-ground around measurement and analytics. For firms that wish to maintain a competitive advantage, basic online reporting is no longer sufficient. The opportunity to make information a true strategic advantage is real. The technologies not only exist, but are mature. They provide the means to create a platform that will use information to leverage all the value of your online properties, that will help create relevant conversations at every stage of a visitor's experience, and that will transform the role of online information. This transformation of information use – from the passive voice of a recorder of what has happened to an active agent directing what should happen – is the endpoint and the outcome of warehousing web analytics data.

About

About the Authors

Gary Angel, President and CTO at Semphonic

Bringing over twenty years of experience in decision support, CRM, and software development, Gary co-founded Semphonic and is president and chief technology officer. He's responsible for leading Semphonic's development of web analytics and SEM decision making tools for web marketing professionals. In addition, he helps companies like WebMD, Intuit, American Express and Charles Schwab maximize their web channel marketing through intelligent use of Enterprise Web Analytics.

Before Semphonic, Gary created and implemented multi-million dollar CRM systems whose customer-driven use of web analytics increased profitability and customer loyalty for Fortune 500 companies including VISA, Bank of America, and American Express. His pioneering work in web behavioral profiling, neural-network web analysis, web-content success correlation, and point-to-point process analysis has enabled major brands such as American Express, AOL, Morgan Stanley and Hotwire to dramatically improve customer acquisition and satisfaction, shopping-cart conversion and overall web channel performance - often with multi-million dollar bottom-line improvements.

Gary has published articles on Web and SEM Analytics in DM News, American Demographics, CRM Guru, CRM Buyer, IMediaConnection, Business Geographics and Business Insurance. He graduated, with honors, from Duke University and lives in San Francisco with his wife and two young girls.

June Dersheiwitz, Vice President of Analytics at Semphonic

June enjoys helping companies improve their practice of web analytics. She held her first web analyst job at a San Francisco startup in 1999 and has worked in the field ever since. Between 2004 and 2007 she took freelance consulting assignments with enterprise clients such as Oracle, CNET, Wells Fargo and Blue Shield. In 2007 she joined the leadership team at Semphonic. She has been actively involved with the Web Analytics Association since its inception in 2005; she currently serves as an elected official on the WAA Board of Directors. As an original instigator of Web Analytics Wednesday, she continues to organize and attend WAW events throughout the San Francisco Bay Area. She has spoken at conferences such as eMetrics, X Change and the Optimost Optimization Summit, and she blogs at <http://june.typepad.com/>.

About Semphonic

Semphonic is one of the world's leading web analytics consulting firms. Founded in 1997, Semphonic is headquartered in San Francisco and with offices in Boston, New York and Washington, D.C. Semphonic specializes in non-transactional sites (be they ad-supported, lead generation, brand awareness, operational, customer support or public service), helping organizations achieve measurable improvement in the performance of their web channel. We support every major web analytics solution from the Omniture Suite to Google Analytics. Our clients include American Express, the BBC, Charles Schwab, Genentech, Intuit, Kohler, the National Cancer Institute, National Geographic, Nokia, Readers Digest, Turner Broadcasting and many other top brands around the world.

SAFE HARBOR

"Safe Harbor" Statement under the U.S. Private Securities Litigation Reform Act of 1995: Certain information contained in this document is forward-looking in nature. Any expectations based on these forward-looking statements are subject to risks and uncertainties and other important factors. These and many other factors could cause delivery of products, features or enhancements to differ materially from expectations based on these forward-looking statements. Netezza does not undertake an obligation to update its forward-looking statements to reflect future events or circumstances.



Netezza Corporation

26 Forest Street, Marlborough, MA 01752

+1 508 382 8200 TEL

+1 508 382 8300 FAX

www.netezza.com

About Netezza

Netezza (NYSE: NZ) is the global leader in data warehouse and analytic appliances that dramatically simplify high-performance analytics across an extended enterprise. Netezza's technology enables organizations to process enormous amounts of captured data at exceptional speed, providing a significant competitive and operational advantage in today's data-intensive industries including digital media, energy, financial services, government, health and life sciences, retail and telecommunications. Netezza is headquartered in Marlborough, Massachusetts and has offices in Northern Virginia, Canada, the United Kingdom, Germany, France, Japan, Korea, Australia and Singapore.

For more information about Netezza, please visit www.netezza.com.