

Industry Brief | Marketing Analytics

CASE STUDY

With speed to market the ultimate competitive edge and a new product introduced every 3.5 seconds, online advertising networks and marketing service providers need fast access to the data they collect daily for their clients—and the ability to perform meaningful analysis on it.

Companies that advertise online need to track how well marketing campaigns are attracting their target audiences. They're also interested in the price of campaigns down to individual banner ads and in cost per thousand (CPM) impressions, cost per click (CPC), or cost per action (CPA).

On the publishing side, service providers are interested in the effectiveness of different-sized ads and physical placement as well as demographic and geolocation-based targeting.

FACING DOWN THE DATA CHALLENGE

Advertising, whether print or online, typically depends on a massive foundation of data for its effectiveness. Companies—or their advertising agencies, for example—must, at a minimum, balance an ad's reach and frequency, number of possible impressions, and CPM against sales results. The ability of the Internet to collect extensive information about potential buyers, their affinities, patterns of influence, the sites they visit, and what they do while there is arguably one of its most significant business contributions. Yet collecting and analyzing this data is far easier said than done. Consider, for example that a single prospect may view the same ad 20 times a day on different sites and then multiply that information by millions of visitors.

Marketing analytics providers depend on the ability to rapidly run complex, ad hoc queries against huge amounts of click stream data, transforming it into high quality, actionable information. Performing routinely scheduled, "canned" queries typically presents them with relatively few problems. However, as the information collected rises exponentially, current data warehouse technology struggles to respond to in-the-moment, business-driven requests. The result is a frustrated business analyst who is unable to retrieve information he or she believes will provide the businesses a competitive advantage. The marketing company, for example, that can best turn data into optimized ad placement, wins.

Most service providers must work effectively with an almost overwhelming volume of data. They're concerned with the inevitable tradeoffs among how much data to collect, its level of granularity, and how long to store it.

On any given day, sites across a typical online ad network host billions of visitors and report millions of events, such as mouse clicks on banner ads or actual purchase decisions. The net result is that the provider is confronted with billions of rows—or terabytes (TB) of data -- that must be stored, managed and made almost instantly available for analysis. In many instances, operational databases are threatened by data overload. Queries that create complex linkages among data tables can slow processing to the point that database administrators (DBAs) must suspend the queries. Alternatively, IT must redesign the warehouse to meet new analytic requirements, a process that make weeks or even months. Clearly this is an unacceptable outcome for a vendor that differentiates itself through its ability to deliver powerful analytics.

Click stream, defined: Like footprints, a click stream is the trail left behind by surfers of the Internet. It is a detailed record of every site and individual page visited, the order in which they were visited, and how long the visitor stayed. In addition, click streams capture social network and newsgroup contacts and even e-mail addresses where users send or receive mail.

SEEKING A BETTER SOLUTION

Seeking a better approach toward overcoming their analytical bottlenecks, many companies have discovered that typical data warehouse solutions are not architected for their needs. Instead, these solutions are labor intensive and expensive, requiring IT to manually partition and index data and continually fine-tune the system. In addition, most solutions are hardware intensive, requiring a significant investment in increased storage and servers for the additional processing power necessary to manipulate large volumes of data.

By contrast, Infobright's data warehouse software was designed from the ground up to be simple to implement, easy to manage and low cost. Simple, easy and low cost – while providing the scalability and performance you need.

Infobright eliminates the typical process of database tuning, data partitioning and index creation. Additionally, Infobright runs on standard off-the-shelf hardware rather than racks of specialized equipment, and because Infobright tightly compresses data (averaging 10:1 raw data size to amount of data in Infobright), significantly less storage is required. Finally, Infobright leverages MySQL as a "wrapper," which means that users can leverage the mature, tested ETL and Business Intelligence (BI) tools they are already using.

WORKING SMARTER, NOT HARDER

Infobright achieves breakthrough speed and performance—particularly for highly complex queries—through its patent-pending Knowledge Grid and integrated Optimizer. Data enters the Infobright and is tightly compressed and stored in "data packs." The Knowledge Grid automatically

creates a highly compact set of metadata, which stores information about the relationship between packs and statistical information about the contents. When a query is initiated, the Infobright Optimizer uses the metadata to intelligently determine which data packs, if any, are to be decompressed to resolve the query. The Knowledge Grid is created on-the-fly and eliminates the need for specialized partitioning of data and indexes, saving significant processing time and accelerating query response time.

Experiencing Stellar Results

Technology that works smarter is what enables Infobright to deliver the results that underpin its customers' successes.

In a proof of concept with a leading digital marketing company that serves more than 1,500 advertisers and 100,000 web publishers, Infobright was chosen for its ability to process complex queries on massive amounts of production data. In addition, the company was favorably impressed by Infobright's seamless integration with MySQL and a total cost of ownership significantly lower than other database solutions.

Putting Infobright to the test, the digital marketer found that it could load an astonishing 3.2 billion rows of data into the Infobright database at an average rate of more than 105 GB per minute. Infobright also compressed the fact tables at a highly efficient ratio of 30:1.

The total cost of ownership savings is significant. The legacy system ran

"After almost 20 years in the industry, I have never seen a product that can resolve queries on three billion rows of data in seconds, and especially not while running on a standard server. Hands down, this is the best price/performance ratio available today."

Mats Johansson
Senior Consultant

Oracle on a \$200,000, high-powered server. Infobright ran on a standard Dell SC1950 1U rack chassis server. The Oracle system required many man-months of a DBAs time to setup and many more per year to tune and maintain. Infobright required none. Finally, the legacy system could not process the queries that Infobright resolved in seconds. While this doesn't affect the TCO, it certainly makes an impact on the bottom-line.

For more information

To learn more about how Infobright can help online advertising networks and other marketing service providers address their data analysis and management challenges, please visit us at www.infobright.com or download a free trial at <http://www.infobright.com/Products/Product-Demo/>.

INFOBRIGHT
Open Source Data Warehousing

HEAD OFFICE:
47 Colborne Street, Suite 403
Toronto, ON, Canada M5E 1P8
Toll Free. 1 877 596 2483
Tel. 416 596 2483
Fax. 416 352 5674

US OFFICE:
5157 Main Street, Suite 202
Downers Grove, IL 60515
Tel. 630 297 4081

EUROPEAN OFFICE:
ul. Krzywickiego 34 pok. 219
02-078 Warszawa, Polska
Tel. 48 22 626 9409