Case study

Big Data solution boosts HP website revenue, value

Near real-time clickstream analytics improves ability to deliver interactive, personalized website experience

Industry
Technology

Objective
Streamline processing of hp.com clickstream data

Approach
Implement Big Data analytics and storage solutions

IT matters
• Solution easily accommodates 101 terabytes of data generated by 37 quarters’ worth of hp.com visitor clicks, which enables business analysts to drive deeper understanding of hp.com visitor behavior
• Queries returned 1,000s times faster, despite fourfold increase in size of data, allowing users to perform more complex, iterative data analysis and enhance hp.com customer experience
• Industry-standard SQL ensures user familiarity, maximizing acceptance and ROI

Business matters
• Faster, more sophisticated clickstream analysis enables hp.com improvements that lead to incremental revenue from website and improved customer satisfaction with hp.com experience
• 60% reduction in storage space enables team to handle 4 to 5 years’ worth of data on same storage platform that previously only handled 4.5 months’ worth; $10 million in future storage capital costs avoided
• Improved financial tracking improves accuracy of rebate forecasts and ability to execute quarter-end financial adjustments, supporting better decision-making and fiscal management processes
• Improved ability to identify and correct issues with website hardware or software, which reduces risks of degraded customer experience and lost sales
• Improved ability to deliver interactive, personalized website experience, which improves sales conversions and drives sales and revenue

“With HP HAVEn, we’ve realized the promise of Big Data analytics. It’s the technology we need to drive revenue and improve customer satisfaction.”

—John Lormand, director, HP.com Technology

Big Data has value, but to realize that value, businesses need to evolve from legacy batch processing technologies to solutions that support real-time interactive analysis. The HP HAVEn Big Data platform offers a state-of-the-art solution that HP has leveraged internally to improve its clickstream analytics capabilities on hp.com, which serves an average of over 1,875,624 unique visitors per day globally.
It is true that “time is money.”

But so is data.

And today’s companies are increasingly aware that the more data they collect, the more value it has. “Big Data”—the enormous data sets generated when companies capture highly granular digital information—can help companies drive innovation and productivity. Big Data can also help companies identify new opportunities and markets, and deepen their understanding of customer needs and behaviors. And Big Data can give companies a competitive edge and help them better understand risk.

But to mine the value of Big Data, companies need cutting-edge technology. They must be able to analyze data sets that dwarf the size of traditional databases. And that analysis must be speedy, to ensure that companies can act on it in a timely fashion.

That is why HP has implemented its own technology, the HP HAVEn platform, to create a robust and comprehensive Big Data analytics solution to improve customer experience on HP’s web properties.

**Billions of clicks**

As is true for many companies today, HP’s public face is its corporate website, hp.com. The site is visited by an average of over 1,875,624 unique visitors per day globally, and functions as one of the company’s most important marketing communications vehicles. The site allows HP to offer thousands of pages of searchable information about its products and services directly to the public. The site also serves as a virtual storefront, enabling HP to engage customers and transact business with them.

During the course of the millions of interactions with site visitors, hp.com generates “clickstream” data, including information on what pages visitors load, how much time they spend on each page, what links they click, and how they exit the site. Analyzing this clickstream data, in turn, allows HP to deepen its understanding of website visitors. The company can better understand how visitors interact with the site. As a result, the data can be used to improve the site itself—making it more usable, for example, or ensuring visitors can easily find the information they need.

More broadly, analyzing clickstream data yields insight into customer behavior, such as buying behaviors. This enables HP to refine its sales and marketing campaigns, or even its products and services themselves. “The biggest consumer of clickstream data is our marketing analysts,” notes John Lormand, director, HP.com Technology. “It is broadly recognized for its value in helping us refine how we communicate to the public and position our solutions.”

At one time, HP stored its clickstream data using traditional Oracle databases, and performed modeling and analytics with SAS Analytics software.

But the data sets have become enormous, due to a number of factors.

First is the volume of hp.com traffic and the number of clicks each visitor makes per visit. “With an average of 531 million page views per month, our data volume is enormous,” Lormand says. To fully support trending and comparative analysis, HP must store around 37 quarters’ worth of clickstream data; analysts typically want to work with about 15 months’ worth at a time to perform year over year trend analysis. This allows the analysts to account for seasonality and show correlation to previous year’s traffic.

The site itself is also extremely complex—more a collection of services than a single application—and is not a static environment. Many pages are generated dynamically, based on information provided by the visitor or visitor behavior. “HP.com is an integrated environment, with some pieces generated by HP and others served up by service providers,” Lormand explains.

HP’s clickstream database, in fact, was HP’s largest Oracle instance.

The sheer volume of the data collected created issues, however. The database performance was sluggish; queries could take days to process. “Query results were taking at least 48 hours after each day’s transactions were completed,” Lormand notes. “And more complex analytics were impossible to do, in practical terms—they simply took too long.

“We knew we needed to improve our clickstream analytics capabilities.”
Big Data analytics, user-friendly model

So HP turned to its own HP Software portfolio and leveraged two components of the HP HAVEn platform.

The first HP HAVEn engine, HP Vertica, is a massively scalable database platform, custom-built for real-time analytics, and sophisticated storage and execution functionality, thereby achieving significant performance improvement in line with industry benchmarking. HP integrated Vertica solution with a second HP HAVEn component, Apache Hadoop, to serve as the solution’s distributed file system.

“Both applications are massively parallel processing systems designed for low-cost Big Data processing,” notes Lormand. “And their functionality is complementary, which means that integrating them delivers substantially more value than would otherwise be possible. We get Hadoop’s ability to load both structured and unstructured data quickly and efficiently. We also get the efficient, extreme analytics supported by Vertica. It adds up to exactly the kind of power we need to handle the volumes of data we collect, and to process it as quickly as we need to process it.”

Since deploying HP HAVEn, the company’s analysts can now get query results exponentially more quickly than they could with their legacy Oracle solution—despite a fourfold growth in the size of the data. With 101 terabytes of data, compressed to 29 terabytes, average query times take less than a minute; queries that previously averaged 24 hours to process can now be completed in 15 seconds—an impressive 5,700x improvement in performance.

Another key advantage delivered by Vertica is that it is based on ANSI SQL, a structure that is familiar to HP analysts. “Vertica offers Big Data analytics capabilities in a user-friendly engagement model,” Lormand explains. “This helped ensure user acceptance of the HP HAVEn technology. As soon as we rolled out the solution, it was embraced by our analysts.”

The HP HAVEn platform also allowed HP to reduce its storage space requirements by 60%, compared to the space required by its legacy Oracle solution. This, in turn, enables HP to manage 4 to 5 years’ worth of data on the same storage hardware that was previously needed for 4.5 months’ worth. And because HP will not have to expand its storage footprint in the future, the company will avoid around $10 million per year in future capital costs.

Faster, more flexible analytics

Today, the HP HAVEn implementation supporting customer analytics on HP’s web properties easily accommodates the billions of rows of clickstream data generated by hp.com visitors—delivering enhanced analytics capabilities to the company’s business users.

HP business analysts can interact with data more flexibly and fluidly. “Our HP HAVEn solution allows more recursive, repetitive types of analysis on our clickstream database,” Lormand notes. “So now, when analysts notice something of interest, they can easily perform iterative queries. This lets them follow a particular train of thought because they don’t have to wait for days between queries. This creates a ‘conversation with the data’ that helps us uncover those hidden insights in the massive amounts of clickstream data.”

Faster and more flexible analytics, in turn, means HP’s understanding of clickstream data is more sophisticated and nuanced. “Now that queries are processed 1,000s of times faster due to the performance of HP HAVEn’s Vertica engine, our business analysts can correlate data points in ways they never could before because our Oracle solution simply couldn’t process the requests,” Lormand explains. For example, analysts can quickly detect if a coupon code has gone viral through social media—the kind of information that is crucial to managing the customer experience as well as planning future promotional campaigns.

The business benefits of these enhanced analytics capabilities will be significant. HP will be better equipped to improve its website functionality and architecture. It can more easily correlate events across its server farms, for example, which will allow it to identify and isolate anomalies that will yield insights into how website functionality is affecting user interactions. “Our HP HAVEn solution gives us a true, end-to-end picture of our environment,” says Lormand. “And because it gives us faster results, we can respond to issues more quickly.”

Fast, flexible analytics unleashes Big Data’s power
HP will be able to better tailor its website interactivity to the needs of individual visitors, delivering a more precise and granular shopping experience. In the past, for example, the site guided visitors to information on the basis of broad categories. If the visitor seemed to fit the profile of a typical retail customer, that visitor would be guided to one set of solutions. Visitors fitting the profile of a home office user would be led to a different subset of products.

But some visitors don’t always fit neatly into these categories. Now, thanks to the insight gained via the HP HAVEn solution, HP can build website functionality that ensures the site responds appropriately to all kinds of visitors. And this, in turn, will enhance visitor satisfaction and improve sales conversion rates: HP estimates this will ultimately increase hp.com incremental revenue. And the overall hp.com customer experience, as measured by periodic surveys to gauge customer satisfaction with the site, will improve as well. “HP HAVEn gives us the capabilities to drive increased value from one of HP’s most valuable corporate assets, our website,” says Lormand. “That value comes in the form of both increased revenue and improved customer experience, and is therefore strongly aligned with HP’s fiscal goals, strategic goals, and brand.”

HP also uses HP HAVEn to analyze channel rebate data, a task that requires matching serial numbers on rebate claims to 135 million rows of shipment data. This has improved HP’s ability to forecast rebates and to execute quarter-end financial adjustments, which in turn supports better decision-making and enhances HP’s ability to manage its fiscal resources.

Given the importance of hp.com to the company’s revenue and brand, it’s likely that managing clickstream data will continue to be an important use of its HP HAVEn solution. “We know that our website must deliver an interactive and personalized experience to our visitors,” Lormand concludes. “It’s a key strategic goal, and HP HAVEn gives us critical capabilities that we need to achieve it.”