

Case study

HP finds big value in HP Vertica big data solution



Time to analyze clickstream data reduced from days to minutes

Industry

Technology

Objective

Streamline processing of hp.com clickstream data

Approach

Implement big data analytics and storage solutions

IT matters

- Solution easily accommodates billions of rows of data generated by hp.com visitor clicks
- Queries returned in minutes instead of days, allowing users to perform more complex, iterative data analysis
- Industry-standard SQL ensures user familiarity, maximizing acceptance and ROI

Business matters

- 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



“Our clickstream implementation of HP Vertica and Apache Hadoop demonstrates the enormous value of these technologies. We expect more and more HP customers will follow suit and adopt the same approach for their big data analytics.”

—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. HP Vertica Analytics and the open source Apache Hadoop software offer a big data solution that HP has leveraged internally to improve its clickstream analytics capabilities.

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 integrated its own technology, the HP Vertica Analytics Platform, with the open-source Apache Hadoop software, to create a robust and comprehensive big data analytics solution.

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 millions of people each month 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 are enormous, due to a number of factors.

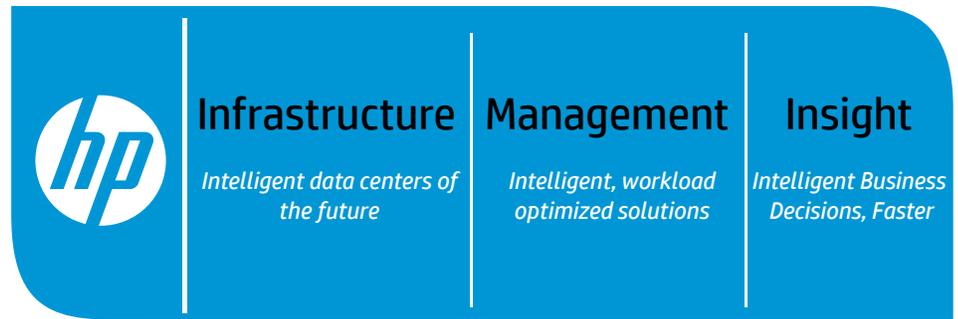
First is the volume of hp.com traffic and the number of clicks each visitor makes per visit. “We capture 11 to 12 billion clicks per month,” Lormand says. To fully support trending and comparative analysis, HP must store around five years’ worth of clickstream data; analysts typically want to work with 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.”



Fast, flexible analytics unleashes big data's power

Big data analytics, user-friendly model

So HP turned to its own HP Software portfolio, leveraging the HP Vertica Analytics Platform, an industry-leading big data solution that supports real-time querying and loading, advanced in-database analytics, and sophisticated storage and execution functionality to speed queries 50 to 1,000 times faster than traditional databases. HP integrated the Vertica solution with Apache Hadoop as its distributed file system. “Both applications are massively parallel processing systems designed for low-cost big data processing,” notes Lormand. “And in terms of functionality, they are highly complementary. Hadoop enables efficient loading of structured and unstructured data. Vertica enables efficient, extreme analytics.”

As a result, instead of waiting days to perform queries, analysts can now get query results in hours or even minutes—even when working with the extremely large data sets that are stored within the hp.com clickstream database.

Another key advantage of 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 technology. As soon as we rolled out the solution, it was embraced by our analysts.”

Faster, more flexible analytics

Today, HP's big data solution—which easily accommodates the billions of rows of clickstream data generated by hp.com

visitors—is delivering enhanced analytics capabilities to the company's business users.

Because the combined HP Vertica and Hadoop solution performs analysis much more quickly, it allows HP to interact with its data more flexibly and fluidly. “Our HP Vertica 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. “Our analysts can correlate data points in ways they never could before because our Oracle solution simply couldn't process the requests,” Lormand explains.

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 Vertica 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.”

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

Customer at a glance

Application

Big data analytics

Software

HP IT Performance Suite—Information Management

- HP Vertica Analytics Platform
- Apache Hadoop

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 Vertica 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 also uses HP Vertica 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.

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 Vertica 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 Vertica gives us critical capabilities that we need to achieve it."

Sign up for updates
hp.com/go/getupdated



Share with colleagues



Rate this document

