CMP
United Business Media

InformationWeek
Business Innovation Powered By Technology
January 1/8, 2007

Exclusive

INSIDE

HP'S DATA WAREHOUSE GAMBLE

By John Foley

FOR MORE THAN A DECADE, BIG COMPANIES and sophisticated data aggregators have adopted data warehouses, yet few have mastered them, and many have outright failed in the effort or have been scared off by the complexity. The goal is to give workers access to real-time data across departments and geographic units, but more often than not, data warehouses end up as costly clunkers with outdated, inconsistent, and missing information.

Hewlett-Packard thinks it can help companies—and not just the biggest and most sophisticated ones—do a lot better. Over the past few months, HP has tiptoed into the

Sacha Lecca

CIO Mott is HP's data warehousing guru —and guinea pig

data warehousing market with a system, called Neoview, born in its research labs and built first for its internal use.

Until sitting down with *InformationWeek* recently, the company has been mum on the initiative—not so much as a peep from its normally talkative marketing team. Indeed, it's an unlikely move into a sector where IBM, Oracle, SAS Institute, and Teradata have years of experience, well regarded products, and loyal customers. Those four vendors—along with Microsoft, which has muscled in on the strength of its SQL Server database—hold about 85% of the $5.2 billion-a-year data warehousing software market, a sector IDC projects will grow 9.5% annually through 2010.

Neoview: The Pieces

- >> Integrated NonStop SQL MX database and operating system microkernel, both modified for data warehousing
- >> Itanium-based Integrity servers in configurations of 16, 32, 64, 128, and 256 nodes
- >> HP StorageWorks storage system
- >> Management dashboard for system monitoring
- >> High-speed data loading and extraction
- >> Support for third-party data analysis tools

Data: Hewlett-Packard

HP's chance of breaking into this established market would seem slim, but a couple of advantages are working in its favor: namely, two of data warehousing's most knowledgeable figures, CEO Mark Hurd and CIO Randy Mott.

Before joining HP in March 2005, Hurd ran NCR and its Teradata division, giving him an insider's understanding of the arcane world of data extraction, algorithms, and database "table joins." Mott has been the country's highest-profile data warehouse user, having built and managed huge Teradata-based systems as Wal-Mart's CIO in the 1990s, then as CIO at Dell earlier this decade.

Teradata in particular seems vulnerable to HP's push. No outsiders know Teradata's products or strategy better than Hurd and Mott, who talk up Neoview when visiting customers—including Teradata users.

EATING ITS DOG FOOD

Neoview's proving ground will be HP's own data centers. HP's IT organization fired up an enterprise data warehouse last May as part of a three-year overhaul of the company's IT operations, put in motion shortly after Mott joined the company in July 2005. That larger project involves reducing the number of applications used within HP from 5,000 to 1,500 and consolidating 85 data centers to six.

The new data warehouse must evolve in lockstep with that wider initiative. It already has 180 terabytes of raw data and 75 terabytes of "usable" data. By 2008, it will be at least twice that size. Some 50,000 employees, a third of HP's global workforce, will have access to it. Eventually, HP's suppliers, distributors, and business customers will too, Mott says. A stickler for deadlines and accountability, Mott is managing HP's three-year IT overhaul and joined-at-the-hip data warehouse with the hands-on attention Ike gave to D-Day.

But Mott and his staff didn't take the safe route in choosing the underlying technology. Rather than go back to the Teradata platform Hurd and Mott know so well, the company is gambling on technology from its acquisition portfolio: the Tandem NonStop operating system and database.

Tandem Computers built a multibillion-dollar business around NonStop in the 1980s and '90s before being acquired by Compaq Computer in 1997, which was itself acquired by HP in 2002. NonStop has a sterling reputation for transaction processing, but it's unproven as the cerebral cortex of a business intelligence environment, where sorting and joining large tables of data require a different feature set (see story, p. 4).

HP engineers had been tinkering with the NonStop software with that in mind before the arrival of Hurd and Mott, but it wasn't until Mott's organization gave its stamp of approval that HP decided to forge ahead with a commercial product. "We had a very strong influence on their overall road map," Mott says. (Full disclosure: Mott is a member of *InformationWeek*'s editorial advisory board, a relationship that had no bearing on this story.)

Neoview was conceived as a data warehousing appliance—similar to those sold by Netezza and others—then morphed into a high-end system. A few pages on HP's Web site describe the product line, though there's been no formal unveiling and HP's strategy remains mostly hearsay in the market. The line consists of the NonStop OS microkernel and database, HP Integrity servers and StorageWorks storage system, a management dashboard for monitoring system performance, and the ability to extract data from line-of-business databases and load it into the warehouse.

That's most of what you need to build a large-scale data warehouse, with one important exception: tools for data analysis. For those, HP is partnering with BI specialists, including Business Objects, Cognos, Hyperion, Informatica, MicroStrategy, and SAS. HP custom developed Java-based BI reporting tools for its internal rollout, but it has no plans to commercialize them. That part of the market could prove hard to resist; IDC pegs the data warehousing tools business at $9.6 billion a year, even bigger than the database portion.

HP has been hiring aggressively to build out the Neoview development team. Greg Battas, chief architect, says his

group has doubled in size and now exceeds 100 database specialists and other software developers. Much of their work has involved writing a database compiler that's efficient at the kind of complex table joins common to data crunching. They've also tuned the system to handle mixed workloads—for example, scanning database tables for analysis while simultaneously processing new data.

The focus now is on creating improved management and monitoring tools for Neoview and "hardening" the system for the rigors of everyday run-your-business analysis. That explains HP's understated approach so far; it doesn't want to overstep its readiness. "It's been made clear to us that you only get one shot at this market," Battas says.

The development activity extends all the way to Beijing. A half dozen researchers in HP Labs China are working with their U.S. counterparts and with computer scientists at China's top universities to write software for moving huge batches of data, says Liu Wei, a research director with HP Labs. The algorithms spread computing loads more evenly across processors.

AMBIDEXTROUS WAREHOUSE

After two-plus years of development work, HP has its first, as-yet-unannounced customer. Retailer Bon-Ton Stores, which operates 272 department stores and seven furniture stores in 23 states, is using a 64-processor, 7-terabyte Neoview system for merchandise analysis and marketing. Bon-Ton has used NonStop systems for transaction processing since the mid '80s and has been running what CIO Jim Lance describes as a first-generation data warehouse on NonStop for 10 years. When HP ran the retailer's data-analysis workload on Neoview, answers came back 13 times faster. "That clinched our decision," Lance says.

Bon-Ton's new data warehouse includes data on merchandise, customers, and suppliers. Other companies have been testing Neoview, and HP promises to announce several customers over the next few weeks.

Here's HP pitch: Data warehouses have fallen short of expectations because the technology has been expensive, proprietary, and used to support only segments of a business, not all data across a company. Neoview will be different. Because Neoview servers have Intel-manufactured Itanium processors at their cores, they're industry standard in design. And HP's rejiggered NonStop software is ambidextrous: It can be used for both real-time data processing and archived data. This so-called mixed workload runs on NonStop's massively parallel architecture, the benefits of which include scalability and five-nines availability.

From a meeting room at a nondescript HP building in Austin, Texas—a one-time Tandem facility that's now home to HP sales, marketing, and technical people, including former Tandem and Compaq employees—Mott laid out the thinking behind HP's internal data warehouse. (A 125,000-square-foot data center is nearing completion next door, one of the six that will serve the company going forward.) Mott says a kick in the pants from the CEO was the impetus for the data warehouse. In his aggressive cost-cutting and realignment campaign, Hurd became frustrated when he couldn't get precise information on the company's global operations from its more than 750 data marts. "There was no lack of data," Mott says. "But there was a lack of consistent, timely data spanning different parts of the business."

Mott knew from experience that an enterprise data warehouse was the answer, and with time of the essence, a Teradata system would have been the easy choice. The Wal-Mart Teradata warehouse he helped build, at 570 terabytes today, is the envy of other companies, and Mott says HP considered Teradata for its data warehouse, as well as a "go-to-market partnership" with the company.

But HP engineers had been developing data warehousing capabilities for NonStop, and Mott needed to give that project a look and determine quickly if HP's in-house technology was ready for wide use. For four months in late 2005, his team ran test loads in the lab. The NonStop-on-Itanium system worked to Mott's satisfaction. Six months later, in May 2006, HP started rolling out the data warehouse internally. Shortly thereafter, Neoview was quietly launched as a commercial product.

Internally, HP is decommissioning its 750-plus data marts at a rate of dozens per month. The data warehouse is being implemented in phases, scheduled for completion in July 2008. When finished, the centralized system will provide hundreds of standard reports and answers to ad hoc queries on everything from product shipments and sales to customer contracts and support calls.

But Mott needs to execute flawlessly. HP's internal data warehouse will serve as a showcase for Neoview; any slip-ups will surely be used against the company's marketing efforts.

InformationWeek.com

MARK HURD talks growth, efficiency, and investment at HP: *informationweek.com/1106/hurd.htm*

CIO RANDY MOTT'S top five IT projects: *informationweek.com/1106/mott.htm*

DATA ANALYSIS from your smartphone: *informationweek.com/1118/smartphone.htm*

WHAT'S NEXT? Predictive analysis: *informationweek.com/1091/predict.htm*

WELCOMING COMMITTEE

If competitors are worried about HP's move into the data warehouse market, they're not showing it. "If HP is going to stir up the pot, we love it," says Teradata VP Randy Lea. With Hurd and Mott and Ann Livermore (chief of HP's technology solutions group) and Scott Stallard (head of HP's enterprise storage and servers division) knocking on doors with Neoview brochures in hand, data warehousing will get exposure among more companies, Lea reasons. And because no CIO is going to award HP a data warehousing contract without evaluating alternatives, Teradata likes its chances.

Revenue at Teradata has been rising gradually, to $378 million in the quarter ended Sept. 30, a 5% increase over the same period a year earlier—not exactly runaway growth, but still a highlight of parent company NCR's performance for the quarter. Last fall, the company released Teradata Warehouse 8.2, the latest version of its flagship software, with 44 new and improved features. One is the ability to partition join indexes for improved performance. An upgraded hardware line is due in the first half of this year. The company has a handful of accounts, such as Bank of America, Cingular Wireless, eBay, and Wal-Mart, with data warehouses in the range of 500 terabytes. Teradata's scalability, unlike Neoview's, is proven.

HP will attempt to portray Teradata as proprietary and expensive, running on computer hardware so specialized that it can't be used for other applications—all of which is true. Richard Winter, a database consultant who worked with HP on its Neoview strategy, expects HP to go after Teradata—and IBM and Oracle, for that matter—on price. HP declined to discuss Neoview pricing, but enterprise data warehouse software, servers, and storage typically start in the hundreds of thousands of dollars and quickly get into the millions. Neoview ships in configurations of 16, 32, 64, 128, and 256 nodes, each with two Itanium processors.

HP isn't a total data warehousing newbie. Many data warehouses run on HP servers, and pros at the vendor's service unit have helped customers deploy more than 1,000 data warehouses over the years. Now, with an integrated software environment on top of its Integrity servers, the company's jumping in with both feet. Realizing it lacked the know-how to be a tier-one player, HP last month agreed to acquire Knightsbridge Solutions Holdings, a 700-employee company that specializes in data warehousing, business intelligence, and data integration services for *Fortune* 500 companies. Terms of the deal weren't disclosed. Gartner, in a written report, describes the Knightsbridge deal as evidence of an HP plan "to build a complete portfolio of BI services, solutions, and products."

Once that acquisition is completed, Knightsbridge will become part of HP's technology solutions group, home to HP's business products and services. Neoview is part of the same group.

InformationWeek contacted IBM and Oracle to get their reaction to the new competition, but neither responded in time for this article.

REDEFINING THE MARKET

The nagging question is, why would HP enter a mature market that's already gone through a wave of consolidation? Informix acquired Red Brick Systems in 1998, then IBM gobbled up Informix in 2001. Mott's answer: As HP aims to redefine data warehousing, "it's not a mature market."

HP contends that most companies have yet to build an all-inclusive "enterprise" data warehouse, and many that tried ended up with systems that serve only portions of their business. Airlines, for instance, build data warehouses for yield management, and telecom carriers to minimize customer at-

NONSTOP KEEPS MOVING ALONG

IT'S BEEN A CIRCUITOUS ROUTE for the NonStop operating system and database. NonStop was developed at Tandem Computers, founded in 1974 by engineers who left Hewlett-Packard. In 1997, Tandem was acquired by Compaq, which was subsequently bought by HP. So NonStop is essentially back where it started.

NonStop's forte always has been high-availability transaction processing, with fault tolerance the main design point of its software and hardware. Tandem's NonStop computers proved popular among banks and stock exchanges for handling escalating transaction volumes with minimal downtime.

In the 1990s, Tandem began rewriting NonStop as a foundation for data warehousing. Through a deal with Microsoft, the NonStop database was built "from the ground up" to run on Windows NT clusters, says Greg Battas, an engineer with Tandem at the time and now chief architect of HP's Neoview data warehouse line, which has NonStop at its core. The NonStop-on-Windows NT systems never made it to market, but Tandem's development team got its first experience working in the world of data analysis.

HP's taking another crack at it. Over the past two years, Battas' team has overhauled the database's compiler, optimized NonStop for mixed workloads of queries and operational data, and created management capabilities for data warehousing. Due next are monitoring tools that are able to predict how system resources get used and adjust accordingly. **—JOHN FOLEY**

HP data center nears completion in Austin, Texas, next to 1980s-era Tandem Computers building

trition. Even Wal-Mart's megasystem contains only a subset of the retailer's data, in areas of supply chain and merchandise, Mott says. A true enterprise data warehouse would contain all that plus information on employees, customer service, and more—in short, 100% of the data a company generates.

It's a radical idea, but not a new one. "The concept has been around for a long time," admits Mott.

Few data warehouse architects have achieved the all-encompassing database for reasons that go beyond the sheer cost and data-integration challenges. With mergers and global expansion, businesses change constantly; a data warehouse that's up to date one day is behind the next. There are turf issues, too. Not all departments want to depend on IT Central for their data-analysis needs. When asked if any of HP's departmental users resisted the mandate of a central data warehouse, Mott responds: "Every one of them."

With its commercial product, HP hopes to enlarge the market, attracting many companies that never considered building a data warehouse by making it simpler and cheaper. That's where the industry standard approach comes in. "Right now, it takes far too much expertise to install and use the systems for data mining,"says database expert Jim Gray, a technical fellow with Microsoft Research who spent 10 years working at Tandem in the 1980s. "There are just not enough gurus to do all the things that can be done."

Borrowing a page from his own playbook, Mott says data warehouses can be self-funding. Eliminate the cost of all those data marts in your company, he says, and you'll free up enough money to pay for a bigger, better system.

But HP still has work to do. Mott admits that Neoview needs fine-tuning in the areas of system monitoring, performance, and integration tools. The gaps in functionality, he says, will be filled quickly.

Consultant Winter thinks the sweet spot for Neoview will be data warehouses up to 100 terabytes, not the 500-terabyte monsters at places like Wal-Mart. Still, systems under 100 terabytes account for 99% of the current market. That's a big enough number to get Mark Hurd's attention.

Mott, meantime, feels a sense of responsibility not only for how well HP's internal data warehouse works, but also for the many other companies he hopes to pull in his wake. Says HP's CIO, "I've got too many friends in the industry to send them down the wrong path."

Copyright© 2007 by **CMP Media LLC**, 600 Community Drive, Manhasset, NY 11030. Reprinted from INFORMATIONWEEK with permission. 7198