MOVING MOUNTAINS

DYNAMIC TECHNOLOGIES ARE ENABLING FINANCIAL SERVICES FIRMS TO MANAGE THEIR CUSTOMERS' ASSETS SAFELY AND AT RECORD SPEED.

ED TAYLOR

Victor Yodaiken, president of FSM Labs Inc.

By Matt Kelly

Anyone who wants to understand the strains on an IT system in the financial services sector need only talk to Pete Abrams, vice president of marketing at security equipment maker NetContinuum Corp.

Abrams can walk a person through the steps necessary to secure a Web-based financial transaction — all 27 of them. In one way or another, financial institutions apply these steps to every transaction they make. And they must manage thousands of transactions every second, executed by dozens of applications running at any moment.

Measure that total demand placed on a financial company's data center, and "you begin to realize," Abrams says, "that there's a lot of computation effort to fully secure and manage every transaction here."

In fact, CIOs and other tech executives these days are configuring their systems to deal with three pervasive issues: hackers prowling for sensitive data; government agencies imposing regulations that require sophisticated IT management systems; and, above all, an unquenchable thirst for faster transaction time.

Little surprise, then, that Abrams and others are banking on the financial sector's eager embrace of dual-core processing. In July, NetContinuum introduced the NC-2000, a security gateway based on Dual-Core AMD Opteron™ processors. Egenera unveiled two blade servers based on Dual-Core AMD Opteron™ processors in April. A host of other AMD partners have also launched Dual-Core AMD Opteron™ processor-based platforms, many looking to satisfy the financial sector's ceaseless need for more processing power.

To alleviate these burdens, CIOs are exploring new IT infrastructures such as virtualized hardware or grid computing, which, in turn, require the agile and high-speed processing enabled by dual-core chips.

"It's a safe bet to say dual-core technology will eventually be much more pervasive at the data center," says Manuel Barbero, vice president with the financial services practice of Virginia-based consulting firm BearingPoint.

"One big usage we see is computing farms," says Egenera founder Vern Brownell, using an industry term for grid computing. Brownell has customers deploying multiprocessor grid systems for calculation-intensive jobs like modeling or risk analysis. "Nearly all the banks I know of have some sort of modeling going on."

Grid computing has long been used to divide and conquer data processing for non-critical applications. Google, for example, employs grid systems to answer search queries because a user doesn't mind when one part of the grid fails and his answers are delayed by a few tenths of a second. In the nonstop financial sector, however, a half-second delay can change the value of a transaction radically.

That timing issue has led many financial companies to restrict their use of grid computing to tasks such as simple

UNDERSTANDING SECTION 404

The heart of Sarbanes-Oxley, Section 404, requires detailed inventories of a company's financial controls and nearly constant testing to ensure those controls work.

What does that mean for financial companies? Even more resources devoted to records retention. E-mails and instant message communication must be archived for seven years. When state or federal regulators come looking for information, they expect prompt retrieval of relevant data.

Moreover, regulators in both the United States and Europe increasingly want to expand the scope of Section 404 — that is, from assessing and controlling risks in financial reporting to assessing and controlling against all risks. For example, the Basel II accords in Europe dictate that a bank must hold a certain amount of cash reserves depending on its risk profile. Ideally, banks will want to calculate that amount — and, implicitly, their risk profiles — automatically.

Such moves toward enterprise risk management are certainly possible, albeit with enormously complex financial models to account for everything from currency fluctuations to housing prices. Processing power, however, has lagged behind user need. Multicore processing will be a crucial tool to meet that new demand.

not vital. An employee could set up a risk scenario, let the computer do the necessary calculations to determine a likely outcome, and repeat as necessary.

"There are many other business problems that can be tackled that way," says Guillermo Kopp, an analyst with the Tower Group, "but the challenge is modeling those problems."

More precisely, the challenge is in modeling those problems quickly. And that's where the AMD Opteron processor can be a big help. With its floating-point engine, financial players can do analyses of far more complex situations. (Pricing the value of an option, for example, requires repeated calculations with randomly generated values to deduce an average price.)

Essentially, says David Jessel, AMD business development manager, one core can run calculations while the other visualizes the data. And because the chip uses HyperTransport™ technology to manage data flowing in and out of the processing cores, the workstations can handle streaming, real-time data while conducting complex analyses.

FSM Labs Inc., an operating systems maker and integrator in Socorro, New Mexico, uses AMD Opteron processors in the operating systems it manufactures. Company president Victor Yodaiken says he historically has pursued customers in fields like aerospace, robotics or instrumentation, but just this August made his first business call to Wall Street — after a year of would-be financial customers pestered him for help.

The biggest driver for multi-core processors, Yodaiken says, is the sheer volume of transactions happening every day. "The volume is so high and response times so short… that people are trying to push this no-touch trading, high-speed data gathering and transactions."

HOW MICROSOFT'S TREASURY DOES IT

With $60 billion in cash and assets to manage, Microsoft Corporation's Treasury Department is a formidable financial-services business unto itself. Microsoft does business across the world, incurring all manner of financial risk from foreign exchange to short-term investments.

The job of Microsoft's Treasury is to manage that risk, maximize the value of Microsoft's holdings and account for expenditures and receipts every day.

In 2003, Microsoft's Treasury decided to overhaul its applications and the IT infrastructure that managed all those financial aspects. The company had been using a 32-bit computing architecture, with insufficient memory to hold both the applications and constantly changing financial data. Those constraints left Microsoft's Treasury analysts unable to run sophisticated modeling programs more than once a week, where Microsoft's vast assets needed updates every day.

The solution: Microsoft's Treasury added four-way and two-way AMD Opteron processor-based servers to its server cluster. The processors lowered the total cost of ownership and the number of system connection points.

The new AMD Opteron processor-based servers ran on Microsoft's public beta version of Windows Server 2003 x64 Edition. "We were going to be the first to use the operating system," recalls Max Giolitti, former group risk manager for Microsoft's Treasury. Giolitti's team simply installed the software on the new machines and flipped the switch.

Instantly, Microsoft's Treasury was able to go from 2 GB to 4 GB of effective memory for the application — enough to hold the software application and data for all of the financial assets. Today, the risk analysis programs that needed 40 hours to run can now do the job in 13 hours.

The implementation "was easy, not difficult," Giolitti says. "There was nothing to it. We didn't have to do any heroics. All we did was take the plunge."

Jeanne Capachin, research director of the Consumer Banking and Credit Advisory Service with Financial Insights, says large banks currently stuck using batch processing overnight to calculate and settle daily transactions will also need more processing power. Small banks, by comparison, with their lower transaction volume, can handle that function in real time.

Egenera's Brownell notes that some CIOs still tend to shy away from innovative IT infrastructures for mission-critical applications such as trade execution or accounting. The "space-age modeling" on the front end, or various other analytic functions, are more suitable candidates for virtualization and data farms, he says.

Still, he adds, the Linux and Windows operating systems already have the muscle to take full advantage of dual-core processing (with Sun Solaris 10 released only in March). He expects "a little bit more maturity" and adoption from software vendors before true 64-bit computing becomes pervasive, but Linux with 32-bit extensions, "works just fine… people can take advantage of that today."

NetContinuum's Abrams agrees that, "the purchase of these concepts is just starting to happen," but says new technologies often take 18 to 24 months to win over the minds of tech leaders. (AMD introduced the AMD Opteron processor in April 2003, and Dual-Core AMD Opteron™ processors in April 2005.)

Manuel Barbero, vice president with the financial services practice of Virginia-based consulting firm BearingPoint, sees a different reason why financial companies will embrace dual-core processing: power consumption. Today's data centers can generate 10,000 watts per rack of servers, or 400 to 500 watts per square foot. "Those densities are not compatible with legacy data centers," which can't handle much more than 35 to 75 watts per square foot.

He expects hedge funds, brokerages and other players in the capital markets to seize on dual-core processing first. Insurers, on the other hand, have been typically late to any new technology and likely to lag the group. Banks move at various speeds, depending on their needs for trading programs or settlement systems.

Regardless of the industry IT adoption trends, Barbero says, "it's a safe bet to say dual-core technology will eventually be much more pervasive at the data center," as IT managers at financial companies try to simplify their complex, overworked, power-hungry data centers.

Gene Banman, chief executive of NetContinuum, agrees. "We have a bunch of best-of-breed specialist servers out there that do one thing very well, and that's great, but once you've gotten five or six of these boxes in your data center, it adds a lot of complexity." AMD

THE TECHNOLOGIES YOU NEED TO KNOW

Grid computing: Also known as "computer farms," grids are an increasingly common infrastructure used to tackle complex computational problems. The grid is a group of dozens or even hundreds of computers, each attacking one small part of a large data-processing chore and quickly re-assembling all the pieces into one answer.

Grids are excellent structures for modeling or data searches, and are already used in scientific and engineering circles. So far, however, grids have been too slow for important financial-services tasks such as securities trading or order management. AMD Opteron processors should alleviate that drag, thanks to their faster I/O interconnects.

Virtualized data centers: IT experts estimate that data centers routinely operate at 20 percent of total capacity, because any one computer can only act as a server or only as a database. No easy way exists to reconfigure boxes into more databases, servers or other equipment as necessary.

Virtualized data centers solve that problem by creating an artificial software world atop each piece of hardware in the data center — one week a box can be a database running on Linux, the next an application server using Windows.

As with grid computing, the challenge for financial services companies will be to find processors that can manage the complex nature of virtualized data centers fast enough for a financial company's needs.

Blade servers: A blade server is a thin circuit board containing several microprocessors and memory, usually intended for a single purpose such as searching a database or transferring files. Shaped like pizza boxes, "blades" conserve space in the data center and let a business pile such servers onto one task for speedy execution.

64-bit x86 computing: AMD pioneered the move from 32-bit x86 computing, which allows only limited access to virtual and physical memory. 64-bit computing, coupled with HyperTransport™ technology, abolishes that bottleneck by putting memory in direct contact with the CPU — resulting in faster applications, more complex graphics, and better management of large data sets.