



GRID TECHNOLOGY

Real Grid

Long the domain of scientists and engineers, grid computing is finally going mainstream. CIOs need to examine their applications to see if their businesses can benefit from grid's power and economies.

BY CHRISTOPHER LINDQUIST

DERIVATIVES CAN be a magic wand for money managers. Used properly, these complex financial contracts help maintain profits by keeping a handle on risk. But pricing them is the key.

For derivative sellers like Wachovia, assessing risk and pricing isn't magic; the software that modeled its derivatives, grinding out the numbers, was complicated and needed to run thousands of what-if scenarios to determine end-of-day prices and to calculate the risk position for the derivatives portfolio. Locked into large, multiprocessor Unix boxes, the risk position calculation could take as long as nine hours. And throwing upgraded hardware at the problem wasn't going to help much. "It would have cut the time from nine hours to four and a half hours," says Mark Cates, chief technology officer for Wachovia's Corporate and Investment Banking group. "We needed it to run in under an hour."

The solution wasn't pricey hardware; it was cheaper hardware. Wachovia linked hundreds of already-deployed desktop computers into a grid, taking advantage of every machine with available processing time. The results were stunning. A job that used to take all day or overnight could now be completed in under an hour, allowing Wachovia to make exponentially faster risk and pricing decisions.

Cates says that the grid solution cost Wachovia a fraction of what it would have cost to upgrade the large Unix environment—an upgrade that wouldn't have produced anything like the same performance benefit. "We're seeing ten- to twentyfold processing increases at 25 percent of the cost," he says.

Wachovia isn't bleeding edge. Thanks to improvements in both hardware and software, numerous companies have begun taking advantage of grid tools. Business users, particularly in the financial services industry, are seeing the benefits of grid in faster responses, reduced time to market for new products, and lower prices per unit of computing horsepower. There are still

hurdles to vault before grid goes mainstream (right now, many apps simply don't make the transition), but grid is no longer just a tool for techies decoding the genome or designing airplane wings.

The Difference Between a Grid and a Cluster

The technology behind grid isn't new. Its roots lie in early distributed computing projects that date back to the 1980s, where scientists would connect multiple workstations to let complex math problems or software compilations take advantage of idle CPUs, dramatically shortening processing times. For years, vendors and IT departments eyed this opportunity to dramatically increase processing power by employing existing resources. But only recently have the tools arrived to put general business applications to work on a grid.

When I first started covering grids two and a half years ago, Sun had defined grids as including clusters," says Joe Clabby, president of technology research company Clabby Analytics and author of a recent report on the state of grid. By that definition, Sun would have had more than 5,000 grids. But while grids and clustering both share resources across multiple machines, grids, according to Clabby, are different because they allow "distributed resource management of heterogeneous systems." In other words, with grids you can quickly add and subtract systems—without regard for location, operating system or normal purpose—as needs dictate. Clusters are built from the ground up to function as a single pool of compute power and consequently aren't as flexible.

The Scale's the Thing

Scaling is one of grid's primary benefits to the enterprise. With properly designed grid-enabled applications, grid can produce staggering performance improvements—add a new processor and get that processor's full power added to the mix. Using grid math, you can add two or more cheaper, slower processors to achieve far greater power than you could with a much more expensive high-end machine. String enough processors together, you can even exceed the number-crunching power of some supercomputers.

Scalability at an affordable price was the key to grid for Acxiom, a company that specializes in cleaning and integrating customer data. Acxiom, for example, can determine if Bob A. Smith and R. Albert Smith in Los Angeles are the same person and, if so, consolidate his customer data into a single record. This is a critical task for marketers looking to maximize the effectiveness of their campaigns, but it takes massive amounts of processing power. Acxiom's "link append engine," called AbiliTec, takes name and address information and uses it to create links to databases, and it does this a lot: 15,000 links every second, 24/7, according to C. Alex Dietz, products leader at Acxiom. "It's a perfect application for grid computing," Dietz says. "Take one name and address and feed it to the appropriate grid node. In

parallel, feed the next name and address to another grid node."

But when Acxiom started integrating AbiliTec into all of its services, the company discovered that the original architecture—based on multi-CPU symmetric multiprocessing machines—wouldn't scale sufficiently to handle the load. Acxiom then built the grid system from scratch using custom code (built around Ascential Software's Orchestrate framework for grid applications) and a bank of IBM blade servers that supply some 4,000 nodes for the grid. (For more on blade servers, see "The Inevitability of Blade" on Page 68.) The result has been the capacity to process 50 billion records a month. "We had to invent a way to take a raw name and address and go into a database and extract links extremely fast and extremely accurately," says Dietz. "We couldn't do it with traditional techniques."

Grid offers scalability in other ways as well. Alain Benoist, CIO for Debt Finance Société Générale Corporate and Investment Banking, says his group moved into grid late last year to help it more rapidly create new financial derivative products as well as to provide faster modeling of the company's market exposure (or "value at risk") for regulatory purposes. But for him, scalability also meant the chance to ease the transition from model to final product.

"The people who are developing the pricing models for derivatives are not people who would feel at home on a supercomputer," Benoist says. Instead, they develop the derivative models using PC-based tools such as spreadsheets. The models are then implemented in common PC programming languages and tested. The ones that pass muster can then be integrated into production applications that can be run on the grid.

But for many companies, scalability and processing power are only half the grid equation. The other half is finding ways to take better advantage of the equipment you already have.

How Your PCs Can Be All That They Can Be

Estimates vary, but the average desktop PC is actually doing something worthwhile less than 20 percent of the time. Some estimates go as low as 5 percent. Yet companies still feel compelled to buy new server hardware capable of dealing not with the average load but with the spikes. Grid promises to solve that problem by putting those idle CPUs to use 40 percent, 50 percent or even 80 percent of the time. And Sunil Joshi, vice president of software technologies and computer resources at Sun Microsystems, claims 98 percent utilization

Joshi doesn't claim that achieving this degree of efficiency is easy, but he does say that through years of practice, his group, which manages computing resources for SPARC processor design at Sun, has turned the maximization of some 10,000 grid-enabled CPUs into a "fine art." His group

even plans the timing of bringing machines down for routine maintenance to minimize any negative utilization impact.

Joshi has an advantage in that many engineering applications have been grid-enabled for years. As a result, his group can tune what applications run where to take maximum advantage of every machine. For instance, a machine could run one application that handles a lot of input and output (I/O) for a database while another, more computation-intensive application hammers the same system's CPU. The goal is to have enough types of applications—high priority and low, CPU-intensive and I/O-intensive—to fill every unused gap, no matter how small, in every grid-connected computer.

Of course, most CIOs would be happy achieving much more modest utilization rates.

"The shortest and best route to getting more utilization, better capacity, is with something like grid," says Philip Cushmaro, CIO and managing director at Credit Suisse First Boston (CSFB).

Cushmaro's organization began using grid computing back in 1999 for overnight batch jobs, work that wasn't particularly time-sensitive and could make good use of otherwise wasted CPU cycles. But as technology improved, CSFB began moving other applications to grid, including critical financial risk management tools. And Cushmaro says the company will investigate other uses for grid. "When everybody goes home at night, all our desktops are doing nothing," he says. "Wouldn't it be nice if we could use those?"

Grid expansion is also on the mind of Debora Horvath, senior vice president and CIO at GE Financial Assurance. Last August, her group began using a grid to run actuarial applications to make financial projections. These computations used to take as long as a day to run a job on a farm of 10 dedicated servers. But by linking 100 desktops (using DataSynapse software) and simply grabbing idle time on the machines, GE Financial Assurance was able to realize performance gains of 10 times over the dedicated (and now turned off) servers without end users noticing anything but the faster response times. Horvath is so happy with the results that her group is already examining new applications that could take advantage of grid. "We have enough other compute-intensive work that we can continue to use grid again and again and again," she says.

What to Say to the Server Huggers

Even with the economies of grid showing promise (including performance gains and the opportunity to take advantage of existing systems rather than buying new hardware), there are still roadblocks to widespread adoption, user resistance among them.

"Whenever you take a significantly different technology to your customers,

they may be skeptical," says Horvath. "When we told people we were going to take away the dedicated servers and use their PCs, there were skeptics. But after we did a pilot, they were won over."

In fact, according to Kevin Gordon, GE Financial Assurance's vice president for IT, new technology and business development, turning the naysayers around took less than half an hour. "We had the actuaries in for a training session. And we took a job that they'd run overnight, and we started it at the beginning of the training session, and within 20 minutes, before the training was over, we had the job completed," he says.

"The politics of grid are a real issue," says Wachovia's Cates. "I believe that's primarily because individual business units lose the ability to control specific hardware." But, he adds, "as the grid approach is proven and use expands, we're hoping the advantages will make that as much of a nonissue as possible."

Standards, Pricing and Other Grid Hurdles

Creating tools that work in distributed, heterogeneous environments is a field ripe for standards, something both grid vendors and customers realize. "Obviously," says Société Générale's Benoist, "if you want to change your applications and move them, you feel better in an open standard world," where applications can work with a variety of grid management software instead of being tied to a single vendor.

Understanding that concern, vendors and researchers are involved in several standards bodies. Key among those are the Global Grid Forum (GGF), the Enterprise Grid Alliance (EGA) and the Globus Alliance. The GGF—whose members include Ascential, DataSynapse, Hewlett-Packard, IBM, Microsoft, Oracle, Platform Computing and Sun—works to develop standards intended to create a wide range of interoperable grid-computing environments and applications. The EGA—formally announced in April by Oracle, HP, Sun and others (though notably not Microsoft, IBM or Platform Computing)—has set goals of providing standards aimed at grid-enabled enterprise applications—what it claims will be a subset of the GGF's work. Globus was formed by a group of research organizations, including Argonne National Laboratory and the University of Chicago and sponsored by the Defense Advanced Research Projects Agency and the National Science Foundation. The group implements standards through its Globus Toolkit, an open-source development suite that lets software makers jump-start their grid development. Current standards include the Open Grid Services Architecture (OGSA), the Open Grid Services Infrastructure (OGSI) and—most recently—the Web Services Resource Framework, which will supplant OGSI later this year, according to GGF, and allow grid software makers to use common Web services standards to identify and utilize grid-computing resources.

Other issues arise around licensing and pricing. Vendors who move their products to grid must figure out ways to price their software. Per-CPU or per-

seat pricing often makes sense in a world where those numbers stay relatively static, but with grid, an application could run on 500 processors one minute and none the next. Being charged for every one of those processors could drive much of the cost benefit out of grid for customers, but adopting a "buy it once, use it everywhere" model could push vendors out of the grid business. Ultimately, per-use price models—likely based on specifications supplied in the OGSA—could dominate, but the tools for tracking such usage have yet to be fully developed.

Grid can even introduce dilemmas for the bean counters. For instance, a CIO at one computer hardware maker says he would love to grid-enable thousands of machines as they go through the burn-in process at the company's manufacturing plants. The machines run various software for extended periods so that the quality assurance people can make sure no components are about to fail before the machine goes out the door. Why not, asks the CIO, run the company's grid-enabled software during the burn-in? "It would be like finding a new source of oil," he says. But the company's accountants can't decide how to describe a product that's been used, no matter how briefly, for work inside the manufacturer. Is it still new? Has it become a depreciable asset? So far, that gusher remains untapped.

Is Grid Right for You?

Right now, many applications simply don't lend themselves to grid computing—especially those that depend more on data handling than CPU power, such as most accounting, CRM and ERP apps. Such applications often take a large chunk of data and run many functions on it, with each task depending upon the previous one. These applications will generally work better on a single processor machine.

The best candidates for grid are applications that run the same or similar computations on thousands or millions of pieces of data, with no single calculation dependent on those that came before. These so-called embarrassingly parallel applications—which include numerous scientific tools, cryptography, and the actuarial and derivative examples mentioned earlier—are ideal for grid, as they scale almost perfectly with the applications able to take advantage of every new processor you throw at them.

Grid standards will undoubtedly help in the effort to move more software to grid by providing a common framework on which applications can be built. And large software companies such as Oracle and SAP already either have products (Oracle 10g) or pilot programs (as SAP does) for grid-enabled applications in place. "Grid has kind of been like a science experiment until now," says Tony Scott, CTO at General Motors. "It's been not commercialized, not standardized, not accessible, not really supported in the classical IT environment." But Scott sees the entrance of Oracle and others as a sign that grid will begin to develop the support infrastructure required by enterprise IT tools: "We not only need the product itself, we need the

manageability tools, we need the provisioning tools, we need all of the ecosystem it takes to support these things in our environment."

It's a Grid, Grid, Grid, GRID World

Grid computing continues to evolve. Analysts and vendors now identify at least three types of grids. While most people think of computational grids, enterprises are looking into data grids that don't share computing power but instead provide a standardized way to swap data internally and externally for data mining and decision support (music sharing systems like LimeWire and Kazaa are examples). Collaborative grids, meanwhile, let dispersed users share and work together on extremely large data sets. (The NEESgrid, www.neesgrid.org, for example, allows earthquake researchers to share data and even research equipment as virtual teams over the Internet.) Clabby of Clabby Analytics also notes that subgenres such as utility grids, enterprise optimization grids and others continue to develop. In short, grid isn't going away.

Given that, it behooves CIOs to identify those business functions and applications that might benefit from being grid-enabled. As GE Financial Assurance's Horvath says, "I think it would be very difficult for a CIO to find a technology and an application that has the payback that [grid] does. The cost is so low and the benefits are so high that it can't be ignored." **CIO**