Enterprise Architect

Print Article

Grid Computing Takes Off in the Enterprise

Start thinking about computing power like you do about electrical power: Pay only for how much you use.

by Mike Ellsworth

Right now, your computer and every other computer in your enterprise is wasting resources. Millions of CPU cycles are doing nothing more than warming the air. Megabytes of disk space sit unused. Your network connection idles as you read downloaded Web pages. On a daily basis, you and everyone you work with typically use only a small fraction of your computers' capacity.

What if you could put all these underemployed resources to work? And what if you could solve the big computing problems your organization has without spending an extra dime on new hardware? That's the promise of grid computing, a method of breaking a big computing problem into smaller chunks and farming them out to dozens or hundreds of computers to work on.

Whether you steal idle cycles from your organization's desktop PCs or link dedicated machines together to form a supercomputer, the result is a lot more computer power to devote to massive computations like modeling weather systems, oil deposits, or viruses; computing complex financial derivatives or insurance risk portfolios; or simulating car crashes. And this computer power is provided with a comparatively small incremental investment—usually miniscule when compared to the cost of buying new hardware.

So what's not to love? You get more resources for your big problems, you get better ROI on your computing investment, and you spend very little. That's why more and more enterprises are looking into adopting grid computing to fill at least some of their computing needs.

Grid is not a panacea, however, and the solutions available are not typically going to help you with scaling Exchange servers or improving online transaction speed, for example. Current grid solutions typically attack more strategic, more massive computing problems, and are not yet suitable for some standard enterprise applications, especially those with a low computing-to-I/O ratio.

Old Concept, New Implementation

Grid computing is not a new concept. In fact, researchers at Xerox's Palo Alto Research Center first explored distributed computing in the 1970s. Along the way, several high-profile grid projects have gained recognition, including the rendering of the *Toy Story* movies through server farm and the cracking—through a grid network in 1997—of Netscape's 56-bit encryption key.

One of the earliest prominent grid computing efforts was the SETI@Home project, which links more than four million Internet-connected computers together into a massive supercomputer. The Search for Extraterrestrial Intelligence project sends each participating computer a chunk of radio telescope data to process. When done, the computers send back the results and request more. Now four years old, SETI@Home has 3.3 million users in 226 countries, has accumulated 1.4 million years of CPU time, is running across 136 operating system/hardware combinations in 226 different countries, and has processed more than 800 million results. The capability of this grid averages more than 48 teraflops per second—not bad for an effort that has cost only a few hundred thousand dollars.

The public model of grid computing seeks to leverage the estimated 10 billion MHz of processing power and 10,000 terabytes of storage connected to the Web. However, this type of public grid computing has never really taken off for enterprise use because of concerns about security and reliability, and for the most part it remains the domain of nonprofit, volunteer efforts. But now—following on the successes of the more familiar cluster-computing concept—grid computing is starting to make an impact in the enterprise.

The concept of grid computing is very similar to the methods used to create many massively parallel supercomputers. The current fastest supercomputers are, after all, federations of thousands of independent processors. The main difference is that supercomputers have much faster connections between processors and between processors and memory than grid computers ever will. A secondary, but perhaps equally important, difference is that in both a supercomputer and a cluster, all the computing resources are under the control of a single organization and are usually dedicated to a single task.

Grid computing differs in that heterogeneous resources controlled by multiple organizations are involved, according to Rajkumar Buyya, program leader of the Gridbus Project at Melbourne University in Australia. "Cluster computing is about resources aggregation in a single administrative domain. Grid computing is about resource sharing and aggregation across multiple domains. In grids, each node has its own resource manager and allocation policy," he writes. Thus, grids can be thought of as loose federations versus clusters' more highly integrated approach. Grid members, for example, are more likely to poll for work to do than to have that work pushed to them by a central authority.

The definition of grid is still evolving, which can lead to considerable confusion in the marketplace. "There will be plenty of marketing-driven confusion in the next few years about what constitutes grids and what problems grid technology can solve," said Kevin Werbach, CEO of the Supernova Group and former editor of Release 1.0 newsletter. "Grids are, in a sense, a further abstraction of computing. Everything becomes a resource, which can be brought to bear on common problems regardless of what platform it runs on or what data center it sits on."

Beyond the Hardware and Network

The readiness of your enterprise to adopt this powerful technology may lie in its willingness to also adopt a new way of paying for computing power. The traditional way to deal with increases, or spikes, in computational needs is to buy more hardware. Grid proponents want enterprises to start thinking about computing power like they do about electrical power: the more they use, the more they pay, but they pay nothing extra for unused extra capacity. "The just-in-time nature of resource utilization in grid computing, a model which enterprises became familiar with during the ASP revolution, offers a significant advantage for organizations that want to maximize limited resources," said Frank Bernhard, a managing principal at Omni Consulting Group. "However, it could be tough sledding getting enterprises to accept this new billing model."

Bernhard points to IBM's "use only what you need" product philosophy that is showing up in its server product lines. You can buy a machine with four processors, and turn off two until the peaks of your business require the others, and then only pay for the extra processors when you need them. "That's the beauty of grid. Its economic viability is determined by the elasticity of a company's computing need." As your need for computing rises and falls, a grid can provide the appropriate resources at a fraction of the incremental cost you would otherwise spend to source and provision new servers (see Figure 1).

Making these grid federations work like a coherent whole on a single task is very challenging. "This *is* rocket science," said Frank Cicio, COO for grid software vendor DataSynapse. "The whole idea is to totally virtualize IT resources. You need to create a dynamic system architecture that behaves the way your business behaves, that can handle the peaks and valleys of demand. You need self-healing—autonomic—systems that maximize your investment in existing resources, taking utilization from an average of 15 percent to 80 percent or more. You need to get beyond the hardware, beyond the network." Using grid, you can do all that while increasing the reliability and availability of your computing resource, Cicio said.

Every application is not a good candidate for grid solutions, and not even every huge computing problem is either. For example, if the solution to a problem requires processors from three different computers, memory from five others and hard disk space from another dozen, the coordination effort and the communications delays caused by even gigabit networking can slow the process disastrously in a grid solution. Such applications are better candidates for massively parallel single-system solutions.

The current best approach to grid computing involves splitting up a very large problem into discrete parts that can be solved independently by a loose federation of computers. Thus, the optimal problems are those that require a low degree of communication and coordination to solve. This "divide and conquer" method requires some sort of central control, in the form of a server that parcels out work, and another that assembles the results. Therefore, grids can be prone to bottlenecks and single points of failure at these critical nodes. Sun Microsystems and others solve this problem by establishing "shadow" master servers that step in if a critical resource fails.

If the average enterprise uses only 15 percent of its IT resources, you're maintaining significant, costly, just-in-case resources to handle computing spikes. "Companies will no longer expand IT systems just because it's possible to," Carl Claunch, Gartner analyst, said recently. Although Claunch believes grids will be best applied in commercial computing applications, "organizations should not assume that the successes in academic and scientific fields can be replicated easily or quickly in the commercial world."

According to a Sun Microsystems' white paper, the types of computing problems best solved by grids include: compute-intensive simulations, often with hundreds or thousands of similar jobs varying only in datasets or parameters; standard daily workloads consisting of many medium-size, single-threaded jobs, each running a few minutes to a few hours, without the need to provide interactive results; and parallel applications with medium scalability (up to 16 processors) and intensive interprocessor communication, or parallel applications with high scalability (greater than 64 processors) and low interprocessor communication.

Premal Kazi, senior product manager for computer vendor Gateway Inc., which announced it will become a grid computing service provider, suggests asking yourself three questions when considering an application for grid computing: Do you have a loosely coupled application, one that it makes sense to parallelize? Do you have a reasonable computation-to-communication ratio? Is it a computation-intensive application?

Using Grid for Business

Many businesses today, including aerospace giant Boeing, financial services company Wachovia Bank, and chipmaker Intel, are using the technology to solve a variety of business problems. In fact, Intel stopped buying mainframes years ago in favor of distributed computing on idle desktops. The company today designs new chips using a system it designed called NetBatch. NetBatch links 10,000 computers worldwide so that when users go home in Israel, their idle equipment is put to use by engineers in California, and vice versa.

Grid software vendor DataSynapse's first application for client Wachovia Bank two years ago used the company's idle processing capacity to cut the processing time for a complex hedge calculation from eight hours to just 15 minutes. Instead of having one price for the whole trading day, DataSynapse's LiveCluster software enabled the bank to price the hedge almost on demand.

Grid's time savings are the most visible benefit from the technology. However, a potentially bigger benefit can be enabling calculations that were not feasible previously. At Wachovia, the company's quantitative analysts were able to develop more complex models containing as many as 100 factors, which were previously impossible to calculate effectively. "We are booking larger, more exotic, and more lucrative trades with more accurate risk taking," said Joe A. Belciglio, the managing director of corporate and investment banking technology at Wachovia.

While the public Internet grid business model hasn't panned out for commercial efforts, Gateway is hoping its new product, Gateway Processing On Demand (GPOD), will be a hit. GPOD is a grid computing service composed of demo PCs in the company's Gateway Country retail stores and built with partner United Devices' technology. The company is targeting small and mid-market companies that don't have the resources to build their own grids. GPOD runs on 6,800 PCs that average 2 GHz, representing more than 11 teraflops of processing power. It also offers a premium service that runs on 120 computers in its data center.

Gateway hopes to leverage United Devices' extensive experience with running nonprofit grid projects over the Internet. "We see there's a lot of value for us in gaining experience with our product in the volatile and difficult environment of an Internet deployment and all the scaling and security concerns that come from that," said David McNett, United Devices' technology evangelist and the cofounder and president of Distributed.net, the first general-purpose distributed computing project on the Internet. "Being able to directly apply that experience to customers' needs is very valuable for us."

According to Gateway's Kazi, the company's advanced engineering team tested the UD software extensively to evaluate the security and unobtrusiveness of the solution before running a customer pilot. He said the UD software gracefully moved aside if a customer stepped up to the demo PC to use it.

Among technology suppliers, IBM has made possibly the largest investment in grid computing. In 2001, IBM announced it was spending more than $4 billion to build 50 computer server farms around the world as part of its plan to turn grid computing into a utility like electricity or water. Earlier this year the company announced ten grid application offerings and is now offering the open-source Globus Project toolset across its entire eServer (AIX and Linux) product line, as well as its zServer mainframe offering. Using zServer, enterprises can add on inexpensive commodity hardware grids to augment mainframe processing.

Aboard the Bandwagon

Other industry heavyweights are also very involved in grids. As you might expect from a company whose slogan is "The Network Is the Computer," Sun is betting heavily on grid computing. The company released its Sun ONE Grid Engine a year and a half ago and claims the software now powers more than 5,000 computing grids worldwide. The Grid Engine extends Sun's significant experience in high-performance computing clustering to the heterogeneous world of the grid.

Intel has been a long-time grid booster and currently links nearly two million PCs around the world in the Intel Philanthropic Peer-to-Peer Program. Microsoft is involved with the multi-institution Globus Project, and is working to hook its .Net initiatives to the grid. Hewlett-Packard also supports the Globus toolkit on its systems, is working with grid software vendor Platform Computing, and has incorporated grid concepts into its Utility Data Center product. Platform Computing provides distributed and grid computing solutions that help enterprises effectively connect, measure, manage and optimize their enterprise resources (see the sidebar, "Logical Steps to Enterprise Grid Computing").

Much of the impetus for grid computing came out of the academic research community. The Globus Project brings together academic computing and commercial computing and its Open Grid Services Architecture (OGSA) specification attempts to standardize the way resources interact on a grid. OGSA brings standards such as XML, Web Services Description Language (WSDL), Universal Description, Discovery, and Integration (UDDI), and Simple Object Access Protocol (SOAP) into Globus, opening the door to Web services-oriented grid applications.

While the Globus standard has the backing of IBM, Microsoft, and Cisco, another methodology, known as Legion, takes an object-oriented approach to grid computing and also has a following, according to Jon Weissman, assistant professor of computer science and investigator at the Digital Technology Center at the University of Minnesota.

Formerly the lead architect on the Legion project, Weissman said, "Globus is more of a bag-of-functions approach, a set of libraries. You can more easily pick and choose which parts you want to use, and thus Globus is perceived as a bit lighter weight than Legion. Globus is going in the direction of everything being a service, while Legion's concept is that everything is an object. But with Legion, you have to buy into the whole model, because security, scheduling, the communications mechanism—everything—is embedded in the object. Consequently, Legion objects turn out to be very big, megabytes in size." The Legion project has spawned a commercial effort, Avaki, founded by Legion's creator Andrew Grimshaw.

Another researcher-oriented industry body, the Global Grid Forum, also is attempting to set standards in this still-developing technology. Then there's the possibility that one of the large multi-institutional projects will spawn de facto standards. Such standards could come out of the TeraGrid project, a multiyear National Science Foundation-sponsored effort to build and deploy the world's largest, fastest, distributed infrastructure for open scientific research; the EUROGRID project sponsored by the European Commission; or the TerraServer, a satellite photo and geographical information service that grew out of a research project sponsored by Microsoft, the U.S. Geological Survey, and Compaq.

Making the Grid Decision

For enterprise architects establishing a grid computing strategy, the key decision may be whether to build or buy. On the one hand, there are various vendor products of varying maturity as well as grid service providers like Gateway to consider. These vendors are hungry for customers, and thus enterprises are likely to have a significant influence in the development of their offerings. This option may lead to quicker implementation times. United Devices' McNett said his company's approach of wrapping native Windows or Linux binaries and trapping their I/O means quick application conversion, taking as little as 10 minutes. McNett claimed 80 percent of applications could be converted using this approach. Once the application is ready to run in the UD environment, there's additional work needed on the front and back ends to split up the work and reassemble the output. McNett said that takes from one to several weeks. "We've actually gone into customer deployments and gotten the app up and running on the first day," he said.

The pricing for vendor solutions is attractive as well. For example, DataSynapse's average implementation is in the $100,000 range. When you compare such a solution to an alternative hardware purchase that might average millions of dollars, an out-of-the-box solution can look very attractive. For enterprises with smaller budgets, Gateway's pay-as-you-go model could be a bargain.

On the other hand, there's the build-it-yourself approach. This approach is perhaps a good choice for organizations that already have parallelized applications and significant IT development resources. IBM is betting that such enterprises will opt for one of their hardware solutions because of their grid support. All the major grid software vendors have APIs you can use, or you can opt for the Open Source Globus, Legion, or free Sun ONE Grid Engine 5.3 toolkits.

Whatever their approach, companies will obviously need to evaluate potential solutions as to their maintainability, performance, security, supportability, stability, scalability, and time to market. Although grid computing has a long history, current offerings are still a bit unproven in the enterprise. Gartner's Claunch said grid computing is "past the 'inflated expectations'" stage of Gartner's Hype Cycle model, which charts the progress of emerging technologies. Grid computing "is climbing the 'slope of enlightenment'" but has not yet reached the model's productivity phase, according to Claunch. (See the sidebar, "Grids Aren't Only for Processing.")

"Grids are a highly custom endeavor today with substantial integration effort required to put one into operation, whether the grid is built fully with toolkits or leverages some productized software components," Claunch said. "In order to mature, grid will need to move to more standardized, off-the-shelf complete solutions of application, middleware, implementation, and management from the one-off [implementations] found today." So if your organization isn't comfortable when the leading edge turns into the bleeding edge, you're probably wise to consider grid computing for small projects first before betting the farm on this next big thing. But however you proceed, it's a safe bet that grid computing will find its way into your enterprise's solution in the near future.

About the Author

Mike Ellsworth is a principal with StratVantage Consulting, a strategic technology advisory firm in St. Louis Park, MN. Previously, he developed B2B e-commerce strategies for an Internet incubator and was a technology manager and architect during a 14-year IT career at ACNielsen. Reach Mike at mellsworth@stratvantage.com.