sageza

Competitive Review
July 2002 Issue 4

Grid Computing: Contender or Pretender?

Part 1: Will computational grids be bigger than the Internet?

By Charles King

On August 8, 1774 the British ship Mariah set ashore a curious cargo at New York harbor. Ann Lee, a religious mystic from Manchester, England, and her eight followers had arrived convinced that the New World would offer their society of "Shakers" relief from the persecution they had suffered in England. However, the Shakers' unconventional religious beliefs and practices made them easy targets for more conventional minds. Lee died in 1784 a year after being brutally attacked by an angry mob, but despite ongoing persecution the Shakers continued to draw converts. In 1787, Lee's successors Joseph Meacham and Lucy Wright gathered the faithful and announced a radical decision: to organize the church into communal "families" whose members would consolidate and equally share their material possessions, ideas, work, and religious worship. By 1794, eleven cooperative Shaker settlements had been established across New England, and in 1805, twenty Shaker villages ranged from Maine to Kentucky, supporting a church membership of about 2,500. By the 1840s, the Shakers reached a peak of nearly 6,000 members. While the Shakers may seem far removed both literally and philosophically from the world of high technology, we believe there are certain parallels between the two that illuminate current and future trends in enterprise computing...

The Sageza Group, Inc.
836 W El Camino Real
Mountain View, CA 94040-2512
650·390·0700 fax 650·649·2302
London +44 (0) 20·7900·2819
Munich +49 (0) 89·4201·7144
Amsterdam +31 (0) 35·588·1546

On August 8, 1774 the British ship Mariah set ashore a curious cargo at New York harbor. Ann Lee, a religious mystic from Manchester, England, and her eight followers had arrived convinced that the New World would offer their society of "Shakers" relief from the persecution they had suffered in England. However, the Shakers' unconventional religious beliefs and practices made them easy targets for more conventional minds, and many in the pre-revolution colonies accused the Shakers of being witches or British spies. Lee died in 1784 a year after being brutally attacked by an angry mob, but despite ongoing persecution the Shakers continued to draw converts. In 1787, Lee's successors Joseph Meacham and Lucy Wright gathered the faithful and announced a radical decision: to organize the church into communal "families" whose members would consolidate and equally share their material possessions, ideas, work, and religious worship. By 1794, eleven cooperative Shaker settlements had been established across New England, and in 1805, twenty Shaker villages ranged from Maine to Kentucky, supporting a church membership of about 2,500. By the 1840s, the Shakers reached a peak of nearly 6,000 members before beginning their slow decline, victims of America's increasing urbanization and industrialization.

What set the Shakers apart from the other utopian experiments that were so popular in early nineteenth-century America was a system of "orders" geared to meet believers' specific needs, as well as leadership organizations designed to maintain communities' spiritual, practical, and financial requirements. Order, in fact, permeated virtually every aspect of rigidly scheduled Shaker life, and was cited by many as the factor that allowed their communities to be so remarkably productive. Additionally, unlike the Anabaptist Amish and Mennonite sects with whom they are sometimes confused, the Shakers enthusiastically employed efficiency-enhancing machinery and other technologies. At a time when the average family farm seldom kept more than 100 acres of land under cultivation, Shaker communities sustained themselves by tending thousands of acres. Shaker workshops, which created the furniture, metalwork, and other implements the sect is best remembered for, easily provided all the goods their communities needed and sold the excess for profit. In essence, the Shakers created communities whose underlying infrastructures were sustained and extended by effectively bringing order to and leveraging the collaborative skills and talents of individual members.

While the Shakers may seem far removed both literally and philosophically from the world of high technology, we believe there are certain parallels between the two that illuminate current and future trends in enterprise computing. In particular, we are struck by the similarities between the effect of grid solutions on enterprise computing environments and the efforts of early Shakers to boost their self-determination and productivity by imposing order among individual members and collaborative communities. This report will discuss the origin and current shape of grid computing, the factors that are influencing its development, and how and why major vendors are integrating grid technologies into their enterprise business offerings.

Grid in the Real World

Anyone who has been ignoring technology news or living off-planet for the past year or so may be understandably confused over the growing fuss about grid computing, which is taking a turn in the media spin cycle as the latest technology to catch some of Silicon Valley's flickering lightning. Advocates claim grid computing has the potential to be as big as or even bigger than the

Internet. While we tend to serve any sort of hyperbole with at least two grains of salt (and prefer it well-roasted, to boot), the characteristics and origin of computational grids are somewhat supportive of this view. From a purely practical standpoint, grid computing might be thought of as distributed computing on steroids, where computing, clustering, and load balancing solutions are shared or parsed out with resource management tools across combinations of networked server and desktop computers for tasks that require large numbers of processing cycles or access to large data sets. If a grid-enabled infrastructure offers consistent, dependable, pervasive access to computing resources, it can be used to provide computational access on demand to widely dispersed end users in much the same way a utility or power grid delivers electricity to businesses and consumers. At this juncture, a wide variety of businesses and organizations have announced support for industry standard grid protocols. Additionally, enterprise vendors including IBM, HP, Intel, Platform Computing, Sun, Microsoft, Entropia, Apple, Avaki, Fujitsu, SGI, and HDS are actively developing or offering grid computing products.

One regularly cited real world example of elementary grid computing is the popular SETI (Search for Extraterrestrial Intelligence)@Home project, where tens of thousands of volunteers leverage unused cycles on computers from PCs to enterprise servers to search radio telescope data for signs of intelligent communication. This peer-to-peer (P2P)-style cooperative effort has delivered more than one million years of CPU time to SETI, impressive by most any measure. However, the SETI project's reliance on the kindness of volunteers is fundamentally different than enterprise grid methodologies. These depend on high level enterprise financial and political support for networking technologies and sophisticated resource management solutions to support what one might think of as the "clusters of clusters" that comprise high-end computational grids combining elements of distributed, parallel, multimedia and collaborative computing processes. This more complex and automated form of grid computing is the model for projects sponsored by NASA, the National Science Foundation, the UK Science Grid, CERN, and the U.S. Department of Energy.

All fine and good, but why should enterprises be interested in or want to buy grid computing capabilities? The simple answers are efficiency and economy. Proponents claim that grid solutions can help businesses more effectively manage and utilize their existing computing resources. To meet increasing computing needs, grid-enabled enterprise processes can be simply or even automatically directed to idle computers or scheduled for slow business periods at night or on weekends. Higher resource utilization can reduce or eliminate the need to purchase new equipment, and increases the business value and ROI of existing hardware. Additionally, grid-based solutions can potentially improve the end-to-end Quality of Service of distributed enterprise applications. Looking out further, grid solutions that extend beyond corporate firewalls could allow companies that partner or collaborate to leverage elements of one another's infrastructures. Business partners could cooperatively share volume visualization systems for R&D projects, or implement data mining applications across complex databases. Even further in the future is what vendors have christened "utility" or "commercial" grid computing: dedicated computational grids designed to deliver computing services on demand to enterprise clients. This "on/off" vision of grid-enabled Web service delivery is what IT evangelists are so excited about.

Grid Evolution

Comparisons of grid computing to the Internet rely on two similarities. First, much of the impetus for and development of grid computing arose from the same fertile government and university laboratory environments that spawned what would eventually become the Internet. Additionally, the intellectual collaboration that lay at the heart of Internet development is alive and well in the grid community. In fact, it could be argued that grid technologies inspire entirely new models of collaboration, since they enable highly complex computing infrastructures to be regarded and treated as singular, interconnected, and interdependent environments. Simply put, if the Internet was driven by the desire to share and leverage information, grids are driven by a desire to share and leverage computational power.

The notion of grid computing began evolving in the late 1980s through the research into running computations across multiple machines that formed the basis of distributed computing. By the mid-1990s, work with Gigabit Testbeds demonstrated the possibility of establishing and maintaining high-speed network connections, and researchers began investigating how to work with complex applications across coherent high-speed networks connecting computers at multiple locations. By the late 1990s, government agencies and universities in the U.S. and elsewhere began programs to network computers at multiple laboratory facilities to support a range of work. The NSF/DOE AccessGrid provides scientists around the world Internet-based collaboration tools including access to lectures and meetings. The Information Power Grid provides computational support for NASA projects including aerospace development and planetary research. More recent projects include the TeraGrid, which will join supercomputer facilities at four U.S. government labs, and the National Digital Mammography Archive, which will centrally store and distribute medical records and data via a dedicated grid to four university hospitals in the U.S. and Canada.

However, grid applications extend both figuratively and literally far beyond North America. The Grid Physics Network supports data analysis for four physics laboratories in the U.S. and Europe. The EuroGrid IST project will establish a European domain-specific grid infrastructure that will connect high-performance computing (HPC) facilities including CSCS, DWD, FZ Jülich, ICM, IDRIS, Parallab, and Manchester Computing. Future projects include the Biomedical Grid, which will link Singapore's biomedical research labs with the country's National University, and the International Virtual DataGrid Laboratory (iVDGL) which will connect HPC facilities in Europe, Australia, Japan, and the U.S.

Even as ambitious, well-publicized projects such as these have been moving forward, a great deal of behind the scenes work is being done to enable grid's future success. The strengths of the grid model rest on interconnecting and bringing order to a wide range of disparate, independent, systems. Its weaknesses stem from the inherent difficulties of making complex, largely heterogeneous systems and computing environments work with to one another successfully. Though custom-built computational grids have been provisioned for several years by IT vendors such as IBM and HP, and ISV/developers like Platform Computing and Entropia, most grid enthusiasts dream of a day when industry standard network protocols will ease the task of developing and deploying truly heterogeneous computational grids. To that end, members of the global grid community formed the Global Grid Forum (GGF). Patterned on the Internet Engineering Task Force, the GGF oversees efforts to ensure the interoperability of emerging grid protocols, and sponsors myriad working groups focused on specific grid issues including security, scheduling, P2P, performance, and architecture. To date, the GGF is supported by over 200 member organizations, including commercial vendors, user groups, and university and government research labs.

Grid Tools and Toolkits

Over time, a number of developer toolkits and technologies have emerged that support grid-style functions or have been used to deploy computational grids. While some of these technologies have been grid-specific from their inception (and evolved from earlier distributed and grid computing efforts), others are more common architectures and protocols that can be applied in grid computing environments.

- Condor — Developed at the University of Wisconsin, Madison, Condor is a workload management system for high throughput computing jobs that provides job queuing, scheduling policy, priority scheme, resource monitoring, and resource management functions. Condor can be used to manage a cluster of dedicated computer nodes or to use idle desktop workstations, and its “flocking" technology allows multiple Condor computer installations to work together across administrative boundaries in grid-style environments. Condor incorporates many emerging Grid-based computing methodologies and protocols and is fully interoperable with Globus solutions (see *Globus Project*).
- CORBA — The Common Object Request Broker Architecture (CORBA) defines some issues that support grid environments, including a standard Interface Definition Language (IDL) for inter-language interoperability and a remote procedure call service, but does not directly address high-performance requirements and specialized devices demanded by grid computing environments. That said, CORBA and grid technologies are essentially complementary, and the GGF sponsors CORBA-related work groups.
- DCOM — Microsoft’s Distributed Component Object Model (DCOM) provides services that are useful in grid environments, including remote procedure call, directory service, and distributed file system, but these solutions do not directly address or affect grid-related issues like heterogeneity or performance.
- Globus Project — Founded in 1996 and centered at USC’s Argonne National Laboratory, the Globus Project is a research and development effort focused on enabling Grid concepts in scientific and engineering computing. To that end, the Project has issued the Globus Toolkit, an open standards-based set of components that can be used independently or together to develop grid applications and programming tools. Additionally, the Globus Project and IBM have proposed the Open Grid Services Architecture (OGSA), integrating grid and Web services concepts and technologies. The Project plans to deliver an OGSA-compliant Globus Toolkit (3.0) over the next twelve to eighteen months. Corporate technology providers including IBM, HP, Microsoft, Compaq, Sun Microsystems, SGI, Entropia, Platform Computing, NEC, Fujitsu, and Hitachi, have publicly announced their support for the Globus Toolkit as an open standard for Grid computing, and several of these vendors are also corporate partners of the Globus Project.

- Java/Jini — Java can be useful for portable, object-oriented application development, but does not address issues that arise in high-performance execution in heterogeneous distributed environments, such as running programs on different types of supercomputers or performing high-speed data transfer across wide area networks. The Globus Toolkit uses Java to provide portable clients, and the GGF sponsors a Jini working group.
- Legion — Begun as a research project at the University of Virginia in 1993, Legion is middleware that can be used to connect networks, workstations, supercomputers, and other computer resources together into metasystems encompassing different architectures, operating systems, and physical locations. Users can draw on these grid-style environments to parallelize complex problems and run programs more efficiently. Dr. Andrew Grimshaw, who directs the Legion project, is also the founder and CTO of AVAKI, a company that develops commercial grid computing solutions.
- UNICORE — UNICORE (UNiform Interface to COmputing REsources) is a European project that is developing access and authentication procedures that will be of particular use in linking HPC platforms and facilities. UNICORE lets the user prepare or modify structured jobs through a graphical user interface on a local UNIX workstation or a Windows PC, then submit, monitor and control jobs through the client. UNICORE provides the underlying support for the EuroGrid IST project.

The Future of Grid

An examination of grid history suggests that these solutions are best considered as simple evolutionary outgrowths of preceding technologies. In other words, the notion of computational grids did not spring fully formed from the forehead of some Zeus-like high tech wunderkind, but instead arose naturally from hard, steady travel on the meandering, intersecting, ultimately converging paths of computing and networking technologies. As such, the development of grid has been and continues to be anything but linear. As can be seen in the wide variety of initial deployments and supporting technologies, grid solutions are flexible enough and proponents are opinionated enough to pursue a staggering number of paths to what remains an essentially singular goal.

What will happen to grid as it inevitably approaches and enters the marketplace? In some ways, we expect continuing complexity to be the norm, at least for the time being. But at the same time, we believe commercial vendors will impose a modicum of discipline on computational grids by developing and delivering recognizable commercial solution models designed for specific grid processes and applications. We will consider the current state of those models, examine how major IT vendors are focusing their grid solutions and strategies, and offer our analysis of the future of computational grids in the second half of this report: **Grid Computing: Contender or Pretender? Part 2: What does it all mean?**