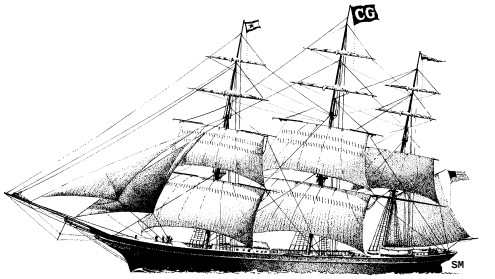
THE CLIPPER GROUP
Observer™



Navigating Information Technology Horizons℠

Published Since 1996 August 26, 2002

Here Today and Growing Fast — What Sun Grids Are Doing for the Enterprise

Analyst: Anne MacFarland

Management Summary

Much of the hype about computational grids paints panoramas of global connectedness and sharing computational resources with others, perhaps even with strangers. It is the 19th century lure-of-new-opportunities-in-the-big-city saga writ cyber and global for the 21st century. A lot of novels have been written about leaving home and hearth for the big city, and global computing could stimulate a lot of movie lines, but epic movies have little to do with Monday morning realities.

There are other grid stories to be told, set in the more prosaic confines of the data center, where order is the ideal and the regularity of business as usual is the norm. In data centers, the increased productivity of employees, particularly the creative workers, has put increasingly strident demands on the data center – not for something new, but for more rapid response time and optimization of existing processes. It is not only capacity that is needed, but more capability from the existing capacity. And here grid provides a solution.

The pain of expensive people's wasted time drove some recently-interviewed Sun customers to install *Sun Grid* software. They were not looking to get an early seat on the future, nor were they analytically placing a technology bet. Instead, they were fixing irritating but prosaic scheduling problems caused by the pace of business, the bulky intransigence of key applications, and their own success.

What these customers found was that grids are surprisingly easy to implement at a local level, and immediately effective. **Like SANs, grids ease mundane but compelling problems of provisioning and scheduling by allowing resources to be pooled.** By their nature, they enable a view of resources that provokes new approaches to analysis of the demand for, and use, of the resource as a whole. Deploying servers in grids is like handling clips of ammunition rather than musket balls. A lot less fumbling is involved, and the results are far more effective. With grids, you move from being the captain of your resources to become general of a capability. And the troops are happier, to boot.

These Sun Grid customers are expanding their grid operations, and expanding their businesses through use of grids. Their experiences give welcome detail to the landscape of what is made possible by a computational grid. In both cases, they implemented a grid to solve a scheduling problem. The implementation then suggested ways to approach business that they would not have contemplated, or considered, without grid experience. For more details, read on.

IN THIS ISSUE

The Experience of Valence Semiconductor

Valence Semiconductor is a designer of residential gateway products. It has major sites in Toronto, Ontario, and Irvine, California, as well as smaller offices in UAE and Russia (St. Petersburg). Valence uses a number of design tools, including *Cadence* and *Mentor*, which are CPU-intensive. These tools involve both interactive workloads and bulky simulation runs. Neither kind of workload can be decomposed for parallel processing, nor can the interactive elements and simulation elements be separately processed. In the semiconductor industry, quick design optimization is a challenge, and quick time-to-market a must. With an engineer-to-CPU of 1.5, Valence needs its servers to be available at all times.

Historically, some engineers had favorite servers, and others would submit the same job to multiple servers in hopes of getting a fast turn around. Neither of these approaches delighted the Valence IT folks. They implemented an internally-developed script to assign jobs to hosts, which solved the first problem, but not the second. Sun Grid "solved" the second problem, because the automated queuing made the old manual submission seem like too much work in a new environment where jobs were running more predictably.

Valence set up a Sun Compute Farm in the third quarter of 2001. All applications are stored in a partition of a very large NAS box that also resides on the grid network, and are pulled to the server as they are needed. They license their software using Globetrotter Software's *FLEXlm*, which allows licenses to be invoked on any of multiple hosts up to the license's stated limit. Valence is also a *SunRay* thin client shop, with only three residual desktops. The corporate culture supports asset sharing.

Selling the grid was no problem. "You get a little test group together and then it grows," said Valence Senior UNIX Administrator David Murillo. Valence started with a group of five engineers at their Irvine facility, added 10 more, and then implemented the grid site-wide. Valence now has independent grids in operation in Toronto and Irvine.

Valence would not consider use of, or participation in, an external grid. The project data is too sensitive. While the grids in Irvine and Toronto are separate and distinct, they both use Tarantella, Inc.'s *Enterprise 3* software with a *SunONE Portal Server* to allow engineers to remotely submit jobs to the Grid. This keeps all the data secure and allows Valence to avoid supporting home computers.

The main choice was between Sun Grid and Platform Computing, Inc.'s *LSF*. In-house development was considered expensive and required a dedicated resource. "Sun Grid had already addressed what we were trying to recreate in a more efficient and robust set of tools," said Mike Susoev, Valence Semiconductor's Vice President of Information Technology. The initial download of *SunGrid* was free, while *LSF* would have cost a substantial amount.

The immediate effects of Valence's grid implementation were to ease their scheduling pain and improve the efficiency of their workload processing. "We learned that pooled resources have a better ROI than dedicated servers per group. Any company that requires multiple processes can benefit from load balancing their resources, not by throwing more resources at the problem," said Susoev. Grid computing allows Valence to purchase and dynamically deploy servers as a universal resource. An additional advantage is the ability to easily reassign resources for a sudden high priority.

The Future of Grids at Valence

With their early grid experience, Valence's IT department sees a need for different auditing and reporting tools. Monitoring a server as a member of a team of assets is different from monitoring a server as an individual machine. Extracting and summarizing performance statistics from the application view, and comparing performance of different servers under different loads, become important. The group dynamics of pooled assets come into play. Larger-scope planning becomes possible, and then, of course, so useful as to be necessary. For Valence, there is no going back.

Mike Susoev feels that more process optimization implementations in the semiconductor business using grid computing will hinge on the tools maturing, both to take advantage of multi-processing opportunities and to modularize, so that CPU-intensive workloads could be separated from user-interactive routines. When the capability of tools permit, Murillo and Susoev foresee setting up a high performance sub-grid for the CPU-intensive routines of simulations runs, and another grid for interactive elements.

Both feel that the potential of grid computing is only constrained by its novelty. Application developers have to buy in and write applications to take advantage of grid deployment. In new installations, heads of departments and small groups need a clear idea of the benefits they will derive from a grid implementation. For a well-focused institution with thin clients and a server farm, that buy-in was not difficult.

Cognigen

Cognigen Corporation provides data analysis for pharmaceutical and allied industries. They initially implemented a Sun Grid (which has grown presently to 22 CPUs) for the ability to queue jobs for *NONMEM*, a FORTRAN application widely used for medical and pharmaceutical data analysis. Developed by the University of California at San Francisco, *NONMEM*, uses 98% of CPU, little memory, no disk, and has no interactive elements at this time.

Cognigen looked also at open source grid solutions, but felt that the alternatives were less functional and harder to implement than the Sun Grid. Like Valence, Cognigen was not looking to enable parallel processing as much as to solve a problem. Scientists were spending too much time chasing down servers to run their workloads.

The Sun Grid installation was a resounding success. Installed in a morning in July 2001, the first test jobs were run on Sun Grid the afternoon of the same day. **With NONMEM and other technical programs, there is a compelling rationale for a single-vendor grid: Floating-point operations must be consistent.** As an existing Sun shop, this was not a problem for Cognigen. They did have more variety than Valence in their end user access devices. The Cognigen end-user environment is two-thirds PCs, and the remaining one-third split between Sun Rays and higher-resolution X Terminals.

Cognigen Grows with Grid

In the process of optimizing existing procedures, Cognigen stumbled onto a new line of business. The installation of the Sun Grid, combined with Cognigen's expertise in software, portals, and high-security environments, led Cognigen to offer Grid-enabled processing to their customers as a portal-enabled utility, in addition to their regular data services. As Darcy Foit, Cognigen's Director of Information Services, put it, "It gives us another thing to talk about with customers." All they had to develop from scratch was an equitable licensing structure, which was no small accomplishment.

The outsourcing mode runs on a separate grid from the in-house grid, and each customer gets a separate portal. Foit anticipates a customer desire for separate grids, and for grids of other platforms. She does not see this as a problem, and even envisions a future involving a herd of sub-grids, including an overflow grid for computational bursts. Sub-grids are also his answer to complex environments, and to the need to control usage of IT resources. Foit also anticipates extending the Grid to other applications, particularly ones like Oracle, for which she feels there are no good clustering options.

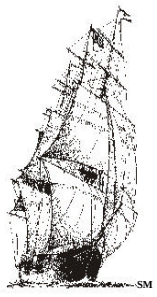
Foit feels that the minimum size for a grid to be economically effective is three or four nodes. This and the free download of the Basic Sun Grid 5.3 (subject to Sun's standard binary license), afford a low cost of entry to a solution that is immediately a welcome alternative to multiple remote mount points.

For Cognigen, the Grid has changed the present and revealed a new vision of future lines of business that will enable faster data analysis and publishing by Cognigen's customers.

Conclusion

These Sun customers tell a counterpoint to the galactic vision of inter-organizational processing grids. They started small, they started local, and they reaped immediate benefits. Their experience continues to give them more ideas of how grids can be used within their organizations. Similar small-scale implementations will allow other enterprises with more externalized business models to get an idea of what an external grid might involve.

Like networked storage, the benefits of a server grid are available at a variety of levels, from modest to broad. Be aware that you won't know the full extent of a server grid's benefits to your enterprise until you try it on for size.



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Anne MacFarland **is Director of Enterprise Systems Research with The Clipper Group.** Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at AnneM@clipper.com or at 781- 235-0085 Ext. 28.***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log,* and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.