The Leading Source for Global News and Information Covering the Ecosystem of High Productivity Computing / February 1, 2008

Features:

Platform Computing CEO Looks to the Year Ahead

by Cheryl Doninger & Charles Coleman, Ph.D
SAS Institute Inc.

Songnian Zhou, CEO of Platform Computing, is widely recognized as a pioneer of grid computing. His Ph.D. thesis established the field of distributed resources management and was the catalyst for the creation of Platform. An accomplished businessman, Songnian has built up Platform from a company of three employees to one that is 400-strong, and has 15 offices around the globe.

In this interview, SAS Institute's Cheryl Doninger asks him his thoughts on how HPC and grid technology are evolving and how that's driving user adoption and vendor opportunities.

Cheryl Doninger: What are your thoughts on how the industry has progressed in the last year?

Songnian Zhou: The HPC industry continued to grow and mature in 2007. This is driven by the maturing and expanding set of compute or data intensive applications. Clusters of commodity computers are ideal for most of these applications to scale out to tackle larger problems, with more accurate results, and in less time. The dramatic reduction in hardware costs compared to proprietary SMP and vector architectures led to rapidly expanding adoption of HPC systems and increase in not only node count, but also vendor revenue. This is a positive feedback loop of a successful market delivering compelling business value.

One sign of progress for this industry is the increasing expectation of integrated systems with not only the hardware but also management software delivered to end users, ready to run their applications. This is in sharp contrast to the "build your own" clusters of 5 or 10 years ago. Another sign is the build-out of technical data centers based on an enterprise grid architecture of integrated clusters. Such centrally managed systems are shared by users across a large organization and managed professionally to ensure availability and low cost.

Doninger: What opportunities and challenges will the industry face in 2008?

Zhou: With the flattening of processor speed and the complexity of multicore, how can the industry continue to meet the insatiable demand for computing power without introducing a lot of complexity? There is no single universal solution. Ironically, just as the industry converges on a set of standard technologies such as x86 processors, Ethernet and InfiniBand switches, and Linux and Windows OSes, it starts to diverge to a variety of technologies such as multicore, GPU, FPGA, and Cell processor. They can be a programming nightmare. There is more need than ever for application middleware and management software to hide the complexity of such technologies from programmers and users while exposing their power. Opportunities and challenges come in equal measure in such a changing market.

An opportunity and challenge beyond programming models is to provide emerging HPC users with simple, easy to deploy and use HPC environments. New HPC users with demands for increased processing power for their compute and data intensive applications are not experts in building and managing clusters, they are looking for turnkey solutions that just work.

Doninger: What should we be doing as (HPC/grid) hardware and software vendors to promote and help facilitate broader adoption of the technology?

Zhou: The components of the HPC stack need to fit and work together to deliver what users care about the most: ease of adoption, ease of use, and low cost of ownership. ISV's need to develop applications and solutions that easily exploit cluster environments. SAS, for example, leads the market for enabling grid and cluster capabilities in their business intelligence and predictive analytics software. This speaks to the increasing value of grid adoption in just about any industry. As the number of business applications that can leverage HPC clusters grows, executives begin to see greater value in the clusters beyond their established space in engineering departments and financial analysis. So, the vendors need to cooperate to ensure their products interoperate. For example, that the applications are certified to run in the clusters users want to use, and the scheduling and cluster management software is fully tested with the OS and hardware. In the end, a cluster should be as simple as an SMP server with OS and management software fully tested and ready for applications. I call such an HPC cluster a "cluster server", as it really should be a low cost and scalable server. Vendor cooperation and product integration is best done based on industry standards and open interfaces; otherwise, the combination of various components to be integrated will explode. The HPC Basic Profile developed with the support of Open Grid Forum is an example of such standard development. The Intel Cluster Ready program is an example of initiatives to ease the certification and adoption of cluster servers.

Doninger: What are the key sectors or industries being most innovative in their adoption of HPC technology?

Zhou: Innovations are coming from a wide variety of industries. For example, the Financial Services industry is leading the way to build internal utility-like enterprise grids operated by central IT providing services to the lines of business. The utility (or shared resource) model opens grid computing up to the entire organization and makes it possible for businesses to start looking at moving a wide variety of applications including business intelligence and analytic applications to the grid. The view from our industry is that the value of the grid scales exponentially as you add more applications to it. In addition to the traditional batch applications for overnight risk management and portfolio evaluation, pre-trade real-time pricing and analytics applications are supported by grid. The application programming model is distributed parallel SOA with many service instances loaded with data and ready to deliver results within milliseconds. In the manufacturing industry, HPC applications are expanding from CAE to manufacturing planning and integrated into the PLM (Product Lifecycle Management) process -- call it HPC automation. With the tremendous power of HPC, business analytics applications are maturing to model and optimize businesses themselves, beyond traditional HPC applications modeling physical products. We continue to see HPC technologies being used in new applications to do things that were not possible before.

Doninger: What technology trends do you see as having the most impact in 2008 and why?

Zhou: Cluster is becoming pervasive, not just for the large enterprises but for users never using HPC in the past. It's the single biggest factor for the growing adoption of HPC. Rather than getting constrained by the power of a desktop or a single server, run it on a cluster server. By sharing the cluster, users get to access even more resources at a lower cost.

Doninger: What effect does consolidation in the industry have on customers?

Zhou: Rational consolidation is a sign of a healthy and maturing industry. Instead of buying software and hardware components from a variety of vendors and putting the systems together, users would rather deal with fewer vendors and get more integrated systems. Consolidation can also lead to economies of scale for the vendors and lower costs for the users. User risks can also be lowered by the increasing viability of a smaller number of vendors. I expect each major part of the HPC systems, such as server, interconnect and management software stack, to be dominated by two or three competing vendors.

Doninger: How will the continued growth in multicore technology affect the HPC industry?

Zhou: The outlook is mixed. It is not clear to me at all whether a wide range of applications will be able to take advantage of the technology possibility of many-cores and squeeze themselves onto single chips. Remember that the popular CPU count for commodity servers stayed at the smallest plural, 2, with some quad-CPU and little beyond. The simplicity of technology, parallel I/O channels of multiple boxes, and independent failure mode of the components have been major attractions of a cluster. I expect this architecture to continue. multicore is a necessary evil the industry has to live with because we simply can't expect processor speed to keep marching up. A cluster of multi-CPU boxes with each CPU containing multi-cores is complex enough for me.

Cheryl Doninger and Charles Coleman, PhD are employees of SAS Institute in Cary, North Carolina, and contribute articles and content in the field of grid, high-performance, and bio-medical computing. The opinions stated here are expressly those of the authors and do not represent the opinions of SAS Institute or Tabor Communications.

Top of Page