

Overcoming Grid Challenges

With rapid changes in hardware and new third-party service offerings, firms looking to deploy and grow their grid computing environments face new and familiar challenges. *DWT* editor Rob Daly sits down with Jingwen Wang, vice president of products with Platform Computing to discuss these major issues.

DWT: With viable utility computing now available, should firms be concerned with rolling their own grid environment?

Jingwen Wang, Platform Computing: One key issue that would seemingly surface is security: data, wire, and most importantly, the intellectual property contained within portfolio models and libraries. Interestingly, it rarely surfaces during one-on-one discussions with customers. With the advent of strong encryption, dedicated lines, and machine cleansing/re-provisioning, the core issue is all about cost. As long as the firm feels it can achieve the same results through utility computing at a lower cost, they are ready and willing to adopt the concept.

It will take time before rented infrastructure can substantially replace in-house, especially for major firms that build competitive advantage through strong internal infrastructure.

DWT: What are the greatest challenges in implementing dynamic resource allocation in a grid environment?

Wang: Technology itself presents the least amount of worry. The main issue is organizational structure and internal politics. However, once you've tamed the "server-huggers," several challenges present themselves. First, there is the issue of ensuring fault tolerance. To ensure one application does not impact others, architectural isolation is required. Specifically, resource and workload management operate independently, ensur-

ing that application issues cannot impact other grid users.

Secondly, there is guaranteeing continued performance and service level agreements (SLAs). Business owners want to maintain and grow performance, and resource sharing implicitly supplies access to more horsepower via access to a larger pool of available power. Sharing guarantees existing performance levels while costing less for additional resources to grow the grid.

Thirdly, there are resource usage reporting and charge-back issues to consider. While sharing enables better utilization and lower costs, flexible tools are required for business owners to take accountability for the use of resources and thus the cost.

Finally, there is the challenge of linking disparate grid technologies. Independent technology silos are spread across application groups and lines of business, creating a management nightmare. The solution to this problem is a common resource management framework with policies to manage any workload scheduler and data management framework.

DWT: How much adoption of cell processors have you seen within firms' grid environments?

Wang: Cell processors and other accelerator technologies are still in the early adoption stage, but interest is starting to build. Evaluators are most interested in the enhanced mathematical processing capabilities, and are taking steps

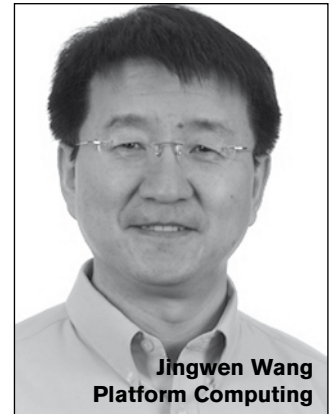
to carve out portions of their application logic that best suit this technology. In many cases—especially where calculation latency is an issue, particularly within intraday processes—the benefits associated with a performance boost far outweigh the complex programming environment.

Accelerator technology vendors like IBM are working to simplify the developer's tasks. In the meantime, grid computing enables "accelerated" business logic to be distributed across multiple cell blades, while the rest of the program's logic runs on traditional x64 hardware. This way, early adopters can focus on the task of optimizing their code for the cell architecture, and leave the distributed scheduling logic up to the infrastructure.

We will start to see more widespread adoption of cell blades once the code optimization exercise becomes more automated. Until that happens, firms that are willing to invest in a more costly onboarding exercise will earn first-mover advantage.

DWT: What is the current limiting factor in growing a firm's grid?

Wang: There are several standard issues to contend with when growing grid environments for traditional high-performance computing (HPC)-type applications, including datacenter costs such as floor space, power, cooling, management costs, and line of business SLA guarantees. The introduction of dynamic resource shar-



Jingwen Wang
Platform Computing

ing and multi-site grids has alleviated some of these concerns; however, firms are looking for ways to leverage their investments even further.

We see many new application classes that traditionally have not been considered high-performance computing (HPC), but in reality will benefit from additional processing power. One example is business intelligence, where firms are now starting to extend resource sharing beyond pricing and risk analysis to customer analytics and even fraud detection.

In many of these use cases, we are seeing a significant move toward the intersection of compute and data grids, where fast and efficient data access has become the bottleneck. Platform Computing's focus is squarely on HPC application environments, and we are relying on partners to enable a data management infrastructure that will drive more "hidden" use cases that leverage the cost-effectiveness of large, enterprise grids, and embrace HPC. We will continue to develop complementary data solutions to address these new use cases as they arise.