



Managing the High-Performance Cloud

by Peter Varhol | Published December 22, 2009



Many of us don't necessarily equate cloud computing with high performance. Cloud computing, where applications run on server clusters, trading CPU time with other applications outside of the enterprise data center, is more associated with traditional IT.

But there are a number of high-performance computing (HPC) sites that operate within a public or private cloud. Many of these are government or university labs that provide computing power to scientists working on research projects. Other HPC sites are affiliated with companies that aggregate supercomputers for use by design engineers within the enterprise.

Private clouds are those that are maintained within an organization entity for the specific use of that entity. It's a data center, but a data center where applications can be loaded and run on an ad-hoc basis. Public clouds, on the other hand, are those where organizations can rent servers to run selected applications. Among the best-known public clouds are Amazon EC2, Microsoft Azure, and Google App Engine.

According to Michael Jackson, co-founder, president, and COO of Adaptive Computing (adaptivecomputing.com), private clouds are much more pervasive today, especially for HPC applications. These clouds manage and execute the analyses, simulations, and other needs of scientists and engineers within the organization.

Making sure that computers in the cloud are fully utilized, and applications are executed efficiently, is a difficult administrative chore. Most cloud applications run packaged in a virtual machine, and can be switched in and out of a server quickly and seamlessly. This offers opportunities for using high-cost computing resources as close to 100 percent of the time as possible, but doing so can be a challenge in practice, especially when the cloud has many

different types of systems, applications, and operating systems.

Dynamic management is especially important for supercomputers, because they tend to be significantly more expensive than generic servers. Not taking full advantage of such computing power is a significant drain on resources and can mean that critical simulations or analyses don't get done.

Adaptive Computing's Moab software manages data centers in a virtualized world, especially if that world exists in the cloud. It offers the ability to manage clusters and grids (both in the data center and in the cloud), and to switch operating systems in and out of supercomputers to get the most effective mix of computing power for the jobs at hand.

Moab Cluster Suite is a policy-based engine that integrates scheduling, managing, monitoring, and reporting of cluster workloads. The grid suite supplements that by managing large numbers of computers arrayed in a grid configuration.

As more and more applications find their way into a cloud for execution, being able to effectively manage the cloud is important to its success as a computing resource. Adaptive Computing's Moab is one way of doing so, and may be the only such product today that focuses on HPC data centers.

Managing supercomputing resources has never been easy, and in the case of HPC, engineers have a large and growing role in doing so. Tools that can analyze workloads and computing mix, and help target how to configure engineering applications on the available systems, will enable engineers and IT to work together to get the most out of the supercomputing data center.

Contributing Editor Peter Varhol covers the HPC and IT beat for *DE*. His expertise is software development, math systems, and systems management. You can reach him at DE-Editors@deskeng.com.

<<http://www.deskeng.com/articles/aaatjg.htm>>