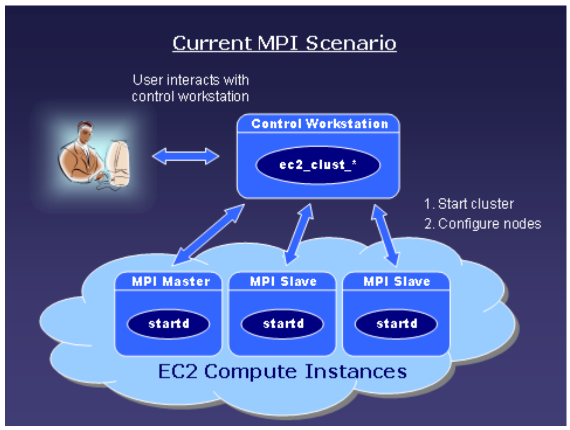
Name: John J. Rehr
Affiliation: University of Washington
E-mail: jjr@uw.edu

Topics: Programming Models for the Cloud
Cloud Self-Monitoring and Autonomic Control

Current Activities

Our current effort is aimed at developing improved tools to enable high-performance scientific computing in the cloud environment, as an alternative to traditional computational resources. This approach is motivated by the availability of these large virtual resources, which have scarcely been exploited by scientists. Our research focuses on the Amazon Web Services (AWS) Elastic Compute Cloud (EC2) and follows a preliminary investigation [J. J. Rehr, F. D. Vila, J. P. Gardner, L. Svec & M. Prange, *Scientific Computing in the Cloud*, CiSE **12**, 34-43 (2010)]. The tools are designed to hide the cloud as much as possible and make it behave like a virtual homogeneous computer cluster – an environment well known to many scientists. Thus, this approach requires no major changes to current computational methodology (see diagram at right). Our initial study of Cloud performance was done with the parallel version of the FEFF code, a widely used x-ray spectroscopy code developed by our group. By comparing the scaling of cpu time vs number of processors between a virtual EC2 cluster and a local cluster, we found that the EC2 does not degrade parallel performance. Recently we have begun porting some advanced scientific codes including WIEN2k and Quantum Espresso, and we are now constructing and benchmarking optimized preconfigured environments for these codes. We are also developing improved user administration tools by linking our cluster tools to the AWS Identity and Access Management (IAM) service. IAM allows us to assign each user an individual access key, and control access to resources on a per-user or group basis. This improves security, tailors the cloud environment to user-specific needs and facilitates management of user-owned and shared data resources.



Future Research Efforts

Building on our initial study outlined above, our future research on "Cloud-Computing-Clusters for Scientific Research" will follow the program outlined in our NSF SI2-SSE grant proposal. To realize this potential, we will develop a complete scientific computing cloud environment (SCCE) that is robust, flexible, easy to use, and cost-effective. Our preliminary results showed that the EC2 can provide convenient access to high-performance clusters for researchers who do not need advanced network performance. Developers also benefit, since our cloud computing approach facilitates the distribution of pre-optimized scientific codes in a controlled, homogeneous environment.

The development of the SCCE follows two complementary tracks:

1) We will provide highly customized virtual clusters built upon images containing pre-optimized applications of interest to the scientific community. Users will be able to access such clusters without need for installation or configuration. These clusters will work intelligently, booting in parallel the required number of nodes, auto-detecting and repairing when virtual nodes experience problems, and scaling resources as needed. Developers can easily add their own applications to our environment, which is highly suitable for scientific computations and benchmarked for various compilers and libraries. Initially, we will focus on important condensed matter and materials science codes: FEFF, WIEN2k, Quantum Espresso, ABINIT, and Exc!ting. These applications will be profiled to provide optimal performance in a cloud environment. In a later phase we will collaborate with developers from more diverse fields. This effort will be facilitated by an AWS grant for EC2 resources.

2) We will improve accessibility by providing a modern graphical front-end built around our cloud cluster toolset. This front-end will require minimal user set-up and will essentially hide the cloud environment, making it as convenient and user-friendly as interfaces currently used to run codes locally. We believe that this effort has the potential to revolutionize high performance scientific computing for many scientists. We expect that a suitable Scientific Cloud Computing Environment (SCCE) will bring high-performance computing to scientists who currently do not have access to suitable resources.