TeraGrid '10 Tutorial Title: Running Parallel Simulations and Enabling Science Gateways with the NSF MATLAB Experimental Resource at Cornell

Length: Half-day tutorial, Monday, August 2, 2010

Level of the material: Primarily introductory.

Agenda: Participants will learn how to run MATLAB programs remotely on the experimental "MATLAB on the TeraGrid" cluster located at the Cornell Center for Advanced Computing (CAC), as well as learn how to access this NSF-funded resource from a Science Gateway or Web portal.

Following a project overview, participants will download and install the client code required to run MATLAB programs on the NSF experimental computing resource. After ensuring that all users are able to run the basic examples provided with the client code, we will walk through the development and testing of a parallel code that will demonstrate important features of the client software. We will also demonstrate how a user can develop an application locally and scale it to the remote resource, and how to debug problems as they arise. Experimentation with the demo code, as well as trying out personal research codes, will be encouraged.

Participants will then learn about nanoHUB, an NSF Science Gateway, and will run MATLAB applications through the nanoHUB portal. Development details will be covered, including how to detect, catch, and diagnose errors; how to run compiled MATLAB codes remotely; how to run non-interactive sessions; and how to run simulations without creating separate TeraGrid accounts for each user.

Instructors:
Dr. Nathaniel Woody, Cornell University Center for Advanced Computing
Dr. Steven Clark, Purdue University Rosen Center for Advanced Computing
Susan Mehringer, Cornell University Center for Advanced Computing

Software requirement: Participants will be encouraged to arrive with a laptop, with the required software (MATLAB 2009a or MATLAB 2009b) and Parallel Computing Toolbox http://www.mathworks.com/products/parallel-computing/ installed. Updates on additional supported versions will be shared with registrants in advance of the workshop. We will have a few laptops available for participants who do not have them. Participants will be provided with accounts allowing them to access the NSF experimental cluster at Cornell.

Prerequisites: Participants should have a working knowledge of MATLAB and a basic understanding of the concepts underlying parallel computing.

Abstract: Cornell University in partnership with Purdue University received a National Science Foundation award to deploy a new experimental computing resource called "MATLAB on the TeraGrid." This cluster provides a parallel MATLAB capability which is available from a MATLAB client running on a user's workstation. Scientists and engineers can seamlessly scale applications from a local machine running Linux, Windows, or Mac OS X to the remote cluster. In addition, Science Gateways technologists may use the cluster as a backend to their portal and integrate MATLAB code into their job submission framework. The pervasiveness of MATLAB in a wide variety of fields has created the need for a computational resource that seamlessly scales applications from the desktop to a larger-scale MATLAB resource without a steep learning curve. MATLAB Distributed Computing Server runs effectively on this resource and provides parallel and distributed computational services to interactive desktop users as well as Science Gateways such as NanoNet, a nanoHUB application. The MATLAB on TeraGrid cluster lowers the barriers to parallel computing for inexperienced users and, at the same time, provides the

more advanced features required by experience parallel programmers. This tutorial will include instruction on how to use the MATLAB on the TeraGrid cluster both as an extension of MATLAB on the desktop and as a simulation tool for Science Gateways. Participants will learn how to seamlessly shift work to a remote cluster operating at CAC as well as how nanoHUB uses this cluster resource to enable hundreds of users to run parallel MATLAB code.