

SEAGrid Use Case - Science Gateways Community Institute Training and Workforce Development

Sudhakar Pamidighantam¹, Linda Bailey Hayden², Eroma Abeysinghe¹

1. Science Gateways Research Center, Indiana University, Bloomington, Indiana

2. Dept. of Mathematics and Computer Science, Elizabeth City State University, Elizabeth City, North Carolina.

Introduction

SEAGrid Science gateway is an infrastructure for computational science and engineering research and education primarily. It integrates computational chemistry application software and computational hardware resources and provides end to end environments with pre- and post-processing tools in a community oriented fashion. The environments are provided as web browser based interfaces and desktop clients where the former provides file upload and downloads required for computations while the latter provides molecular editors and graphical application input generation interfaces and output parsers to extract output components for visualization. The infrastructure is extensible such that other open tools can be integrated tightly or loosely. Such tools have been used in the classrooms for project oriented computational chemistry topics at undergraduate level. Here we discuss recent use of SEAGrid infrastructure for training and workforce development for computational chemistry, focused at Minority serving institutions.

Training

Training new users toward an understanding and executing computational models requires both providing the concepts in the model as well as the practical execution of the model and extracting the results with visual analysis. The concepts involved may require notes on model descriptions, parameter selection/setting and their significance and quality of the results and discussion about the model and its utility. Once the concepts are understood, a practical way to execute a physical model of the system using computations will be needed. A part of this exercise is to build a representative of the systems of interest, and defining parameters required for the physical modeling of interactions that can be consumed by an application software. This is followed by creating a workflow to supply the required inputs by staging data files of the models, parameters, instruction inputs and executing the application where it is deployed. The application execution training then involves using an infrastructure to accomplish these two critical components in an organized fashion to slowly build the confidence of the trainee to be able to explore other similar systems that may not be covered during the training.

SEAGrid gateway has been instrumented to provide the latter part for training in Protein Ligand Docking, Ab initio quantum chemical computational chemistry using multiple applications and Molecular Dynamics. The input preparation step uses molecular structure searches using

Cambridge Structural Database (CSD) and exporting a model from the search into the molecular editor for further manipulation which in turn is exported to a graphical input generator for a specific application software. The inputs are specific to an application software and the interfaces would automatically provide a selection of resources where the particular application is deployed for the users to choose from. The choice of resource to execute the application is typically made with experience as to which resource is likely to return the results at the earliest or a way to distribute the computational if many were to be executed as part of a larger project.

Workforce development

To understand the impact that SEAGrid gateway could have on their research, SGCI conducted a three-day Computational Chemistry workshop for graduate students and included an opportunity to move a research project onto the grid. The workshop took place Tuesday, May 9 through Thursday, May 11, 2017 at Jackson State University in Jackson, Mississippi. The workshop incorporated the Science and Engineering Applications Grid (SEAGrid) interface and empowered researchers to utilize scientific applications deployed across a wide range of supercomputers, campus clusters, and computing clouds.

Multiple Universities and agencies were represented at the workshop. Students at various levels took part in hands-on sessions and presented their research as well. Highlights can be viewed at <http://nia.ecsu.edu/sgci/2017/170509chem/http://nia.ecsu.edu/sgci/2017/170509chem/>.

UNIVERSITIES REPRESENTED

Auburn University
Binghamton University
University of Houston
Jackson State University
Louisiana State University
North Carolina Agricultural and Technical State University
University of California - San Diego

- 18 of the 20 participants completed the workshop evaluation.
- 72.2% male; 27.8% female
- 100% graduate students
- 50% ethnic minorities
- 11.2% indicated they had a working knowledge of SEAGrid prior to the workshop. That changed to 89.9% after the workshop.
- Over 80% indicated that their future careers would involve computational chemistry.

Training assistance at the hosting institution

SEAGrid gateway is used by scholars, professors for their specific institutional training webinars, conferences and workshops. When such sessions take places SEAGrid support team provides several services to the training organizers in addition to conducting training events. These include 1. Configuring training application if it not already available, 2. Provide quick training for the instructors 3. help in creating training materials 4. Assist with hands on sessions and 5. Provide announcement for the training sessions and host the program and training materials online. These are elaborated further below.

Initial discussion with the organizations in order to make the required applications configuration and testing. This phase is carried out by the SEAGrid gateway admins/PIs. Once the applications are configured and tested they would be handed over to the organizers. Once the applications are available we provide training and demo the gateway features for the trainers. They would be the primary tutorial presenter at the training session and we would make sure they are familiar and comfortable working through the gateway features. These training is provided in form of demos, documentation and actual hands on sessions. The methods differ based on the trainer's familiarity with the gateway. For trainings with SEAGrid gateway, the required materials are made available along with any data files needed in the SEAGrid wiki space. This helps users and trainers as everything being used is in one location. For the actual training session or hands on session SEAGrid team supports by providing required any test login accounts, test input data materials and other related material needed. Providing test gateway accounts saves time in creating these by the users during the training event. SEAGrid gateway will also announce and publish any trainings that will take place using the gateway. This will help the training sessions as gateway users will get to know about them in advance. For trainings which are available both in location and online these announcements are a positive way of publishing and broadcasting for any new attendees.

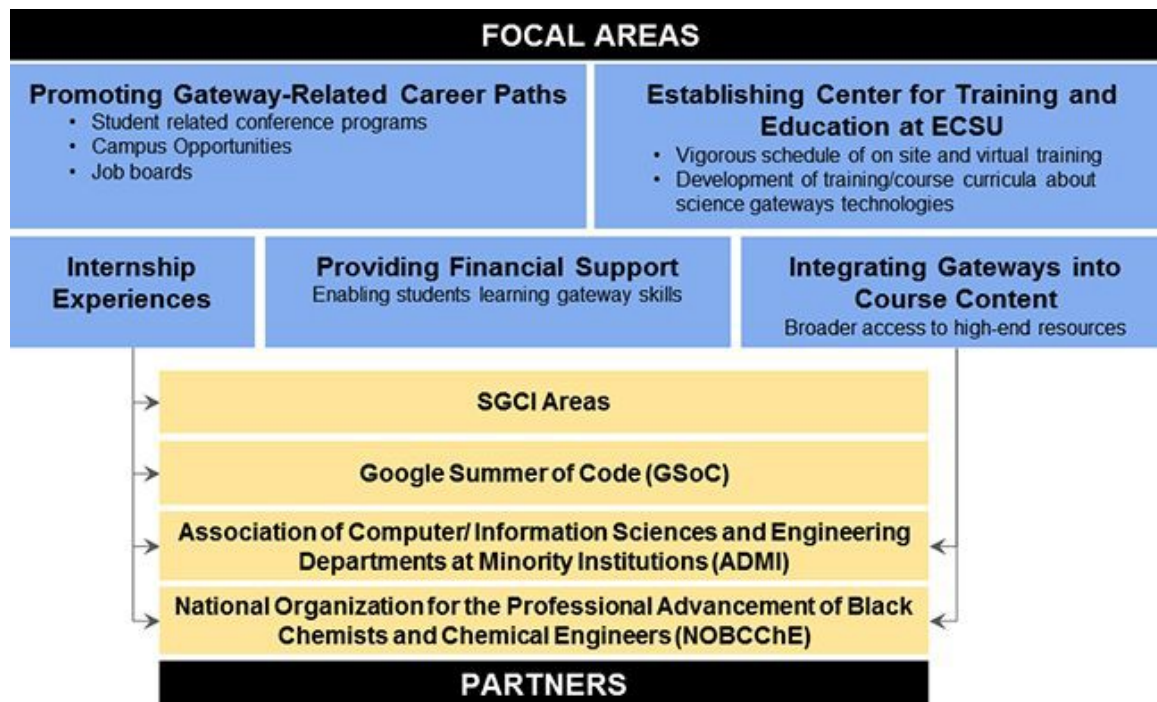
Online Tutorials

Training materials used during these workshops and training events are preserved and presented as online tutorials for use for those who need additional time for review during a training event as well for others who may want to use them for training at their own time. Several of these materials are posted along with the input examples used and detailed stepwise tutorials on the SEAGrid wiki which is accessible from the SEAGrid "Documentation" link.

Mentoring Contributors: In addition to these materials which train users in the domain specific content, SEAGrid development team is committed to extending support for contributors to the various software infrastructure as well. Such mentoring occurs in hack-a-thons and individual interaction sessions for users to register their own allocations on resources such as XSEDE or implement improvements in pre- or post- processing tools for one scientific application or the other.

Broader goals

Workforce Development aims to provide training and professional development opportunities for the next generation of science gateway users and developers while tapping the unrealized potential of students from underrepresented groups. To do this, we are using several strategies:



Outlook

The recently concluded workforce development event at Jackson State University provided us critical feedback in terms of the what the students at graduate level would like driven by their immediate needs toward graduation or post-graduation requirements. A strong desire for extended workforce development sessions interacting with the instructor and more indepth engagement to deploy application software immediately needed, perform advanced computations and drive enhanced capabilities of the gateway itself in terms of pre- and post- processing extensions. We continue to showcase the SEAGrid gateway, both the web browser and desktop interfaces as an exemplar gateway built using for Apache Airavata with a SciGaP hosted middleware and develop new tutorials, workshops and workforce development events.

