928.18
GGG24

A Neuroscience Gateway for High Performance Computing

*N.T. Carnevale[1], S. Sivagnanam[2], K.K. Yoshimoto[2], V. Astakhov[2], A.E. Bandrowski[2], M.E. Martone[2], A. Majumdar[2]
[1]Yale University School of Medicine, New Haven, CT; [2]University of California, San Diego, CA

We are developing a Neuroscience Gateway (NSG) that will provide neuroscientists a user-friendly Web-based interface for using high-performance computing (HPC) and other cyberinfrastructure (CI) resources. This project is motivated by the emergence of research questions requiring simulations that impose a large computational burden and/or heavy storage requirements.

The problem and its solution

Powerful domain-specific open source tools, such as NEURON, GENESIS3, MOOSE, NEST, and PyNN, can be used to construct models suitable for simulation on parallel computers, but few investigators have been able to make use of parallel hardware, and even fewer have run simulations on extreme scale HPC machines. The main reason for this is the high barrier to entry: investigators must divert effort from more productive activities in order to deal with time consuming, confusing details that differ significantly from one facility to another, such as

- writing a successful request for computer time
- mastering complex administrative policies and batch system details
- optimally installing applications on HPC resources
- managing workflow that involves multiple remote authentication schemes
- handling data transfer, storage, and output retrieval issues

The NSG will enable neuroscientists to access HPC resources, funded by NSF and other support, through a convenient interface to simulation software that is already configured for optimal use. To this end, we will build an infrastructure layer, called a science gateway [Wilkins-Diehr 2008], that reduces administrative complexities and abstracts away most technical details. Investigators will be able to easily access the software they need on very large HPC resources through a point-and-click browser environment for configuring and running domain-specific programs. All CI/HPC-related complexities will be hidden within the black box of the gateway architecture. This approach has been successful in other research domains, where gateways such as CIPRES [CIPRES, Miller 2010] (Figure 1) have catalyzed wider usage of HPC resources.

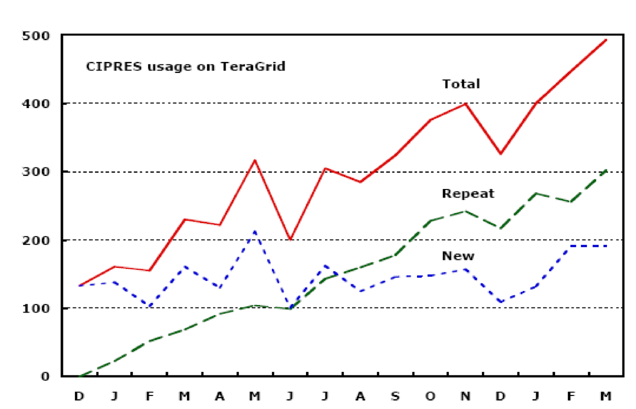

Figure 1. Number of CIPRES Gateway users per month running jobs on the TeraGrid. CIPRES was developed at the San Diego Supercomputer Center (SDSC) for the computational phylogenetics research community. Users/month increased from 132 in 12/2009 to more than 500 by 3/2011. The number of new users/month was >100 from the inception of CIPRES, and the number of repeat users increased steadily.

Design and implementation

The design goal of the NSG is to make it easier for users to

- use HPC time under reasonable usage guidelines and policies
- specify parallel simulation parameters, e.g. number of cores, memory per core/node, estimated job run time etc.
- query job status, and request automatic notification of job completion
- securely access and download output results

The PI and Co-PIs of this project will obtain community allocations on HPC clusters and other resources through NSF and other organizations. Users may employ this community account to submit jobs, and those who have their own allocations will also be able to use their time via the NSG.

The NSG will provide a web portal based user environment for uploading model source code and running simulations (Figs. 2 and 4). It will offer a selection of commonly used neural simulators that are optimally installed on various HPC resources. It will distribute jobs transparently across those resources while providing elegant recovery from failures and minimizing job loss due to system/hardware errors.

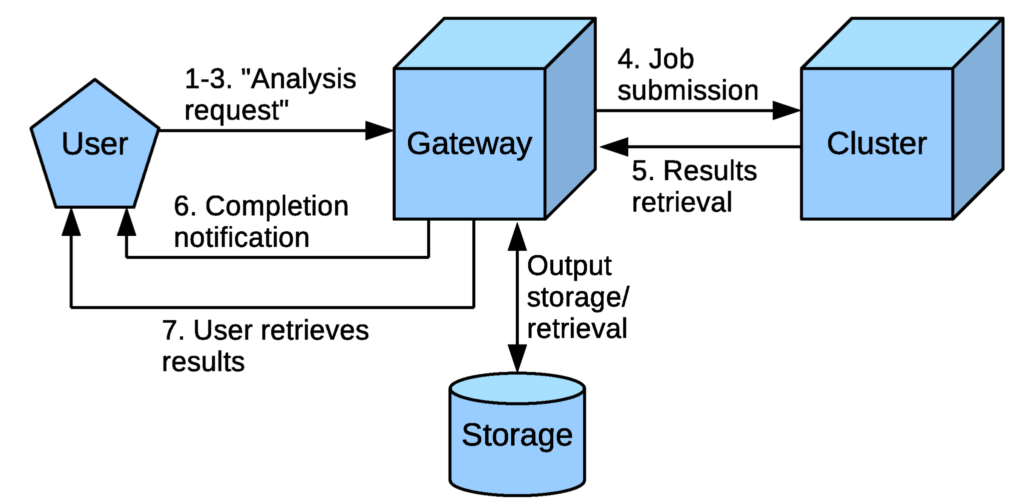

Figure 2. From the user's point of view, the NSG workflow will be simple.

The NSG will also provide

- documentation and examples of how to parallelize neural models for running on HPC machines
- programmatic interfaces to other neuroscience tools (e.g. ModelDB, WBC, NIF, neuroConstruct)
- interfaces for simulator/tool developers that simplify installation and testing of community codes
- a community environment for collaboration, data sharing, etc.

Technical details

The NSG will be based on the CIPRES Science Gateway framework [CIPRES, Miller 2010]. using the Workbench Framework (WF) (Figure 3).

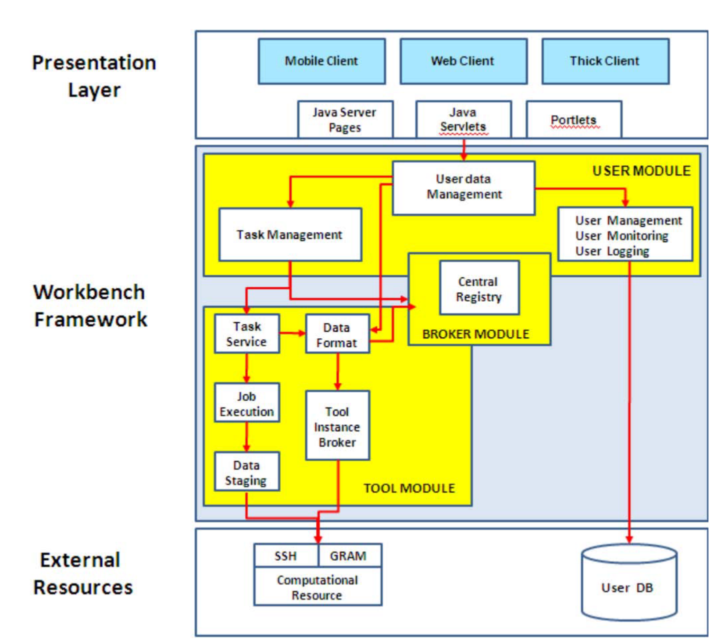

Figure 3. The NSG's WF architecture. WF is a software development kit that contains modules to manage submission of jobs to analytical tools on computational resources, and modules to manage queries on data resources.
Presentation Layer: Supports access by browser clients and programmatic access via SOAP [SOAP] and ReST [Fielding 2000] services.
User Module: Manages all user-initiated activities, passing queries and tasks to executive portions of the infrastructure; stores user and task information; manages user roles, account assignments, data sharing between accounts, access to tools and data sources (which may be proprietary).
Broker Module: Stores and provides access to application-specific information in a Central Registry, e.g. input and output data types for each application.
Tool Module: Translates user-submitted tasks into command lines, submits commands and user data to appropriate compute engine for execution.
External Resources: Can be accessed via SSH, GRAM/Globus [GLOBUS], SOAP, ReST etc..

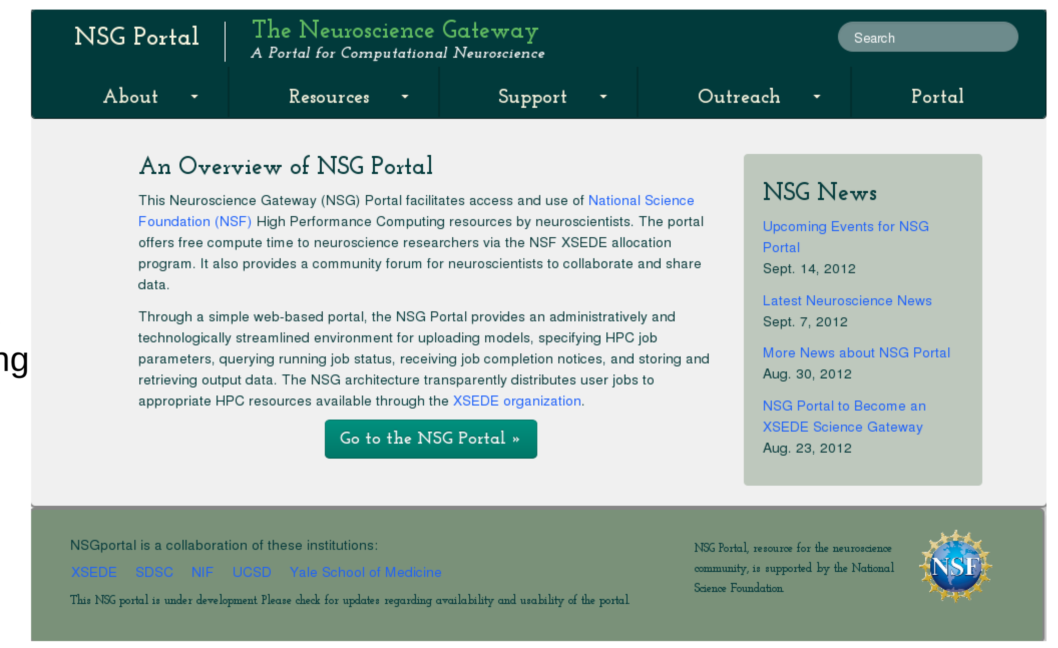

Figure 4. Home page of the NSG Portal: http://www.nsgportal.org/. Initial tests are now being done to prepare an alpha version for testing by neuroscientists in the near future.

Benefits

The Neuroscience Gateway will benefit the broader neuroscience research community in several ways. Chief among these are:

- Its simple interface will provide streamlined access to HPC resourses without forcing investigators to divert effort from their research.
- Users with limited local (university-scale) resources will be able to address pressing research questions that require access to large scale, advanced systems.
- Community code developers (i.e. developers of software such as NEURON, GENESIS3, MOOSE, NEST, PyNN) will be able to test/debug, benchmark, and scale codes on large scale resources and, when satisfied, make them available to the broader user community.
- The NSG can also serve as a vehicle for making new community codes available to users.
- It can be used for teaching classes, workshops, and tutorials without having to set up codes on HPC resources or create new accounts for students/participants.

References

[CIPRES] http://www.phylo.org/
[Fielding 2000] [REST] R. T. Fielding, “Architectural Styles and the Design of Network-based Software Architectures, “ Ph.D thesis, Information and Computer Science, UC Irvine, 2000. http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm
[GENESIS3] http://www.genesis-sim.org/GENESIS/G3/
[GLOBUS] http://globus.org/
[GRAM] http://dev.globus.org/wiki/GRAM
[Miller 2010] M. Miller, W. Pfeiffer, T. Schwartz, “Creating the CIPRES Science Gateway for inference of large phylogenetic trees,” Gatways Computing Environmens Workshop (GCE), 2010, pp 1-8, New Orleans, LA, 14 Nov., 2010. DOI 10.1109/GCE.2010.5676129.
[ModelDB] http://senselab.med.yale.edu/modeldb/
[MOOSE] http://moose.sourceforge.net/
[NEST] http://www.nest-initiative.unifreiburg.de/index.php/Software:About_NEST
[neuroConstruct] http://www.neuroconstruct.org/
[NEURON] http://www.neuron.yale.edu/
[NIF] http://neuinfo.org/
[PyNN] http://neuralensemble.org/trac/PyNN/
[SOAP] http://www.w3.org/TR/soap/
[WBC] http://wholebraincatalog.org/
[WF] http://www.ngbw.org/wbframework/
[Wilkins-Diehr 2008] N. Wilkins-Diehr, D. Gannon, G. Klimeck, S. Oster, S. Pamidighantam, "TeraGrid Science Gateways and Their Impact on Science", IEEE Computer, Volume 41, Number 11 (November, 2008), pages 32-41.