

## The DOE Science Grid

*Project Principal Investigators: William T.C. Kramer, National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory, Al Geist, Oak Ridge National Laboratory, Keith R. Jackson, Lawrence Berkeley National Laboratory, Jennifer M. Schopf, Argonne National Laboratory, and Scott Studham, Pacific Northwest National Laboratory;*

*Management Council: PIs, R. Bair, ANL, I. Foster, ANL, W. Johnston, LBNL;*

*Science Grid Engineering Working Group: K. Jackson, Chair, LBNL, S. Chan, NERSC, K. Chanchio, ORNL, D. Cowley, PNNL, T. Genovese, ESnet, M. Helm, ESnet, D. Olson, LBNL*

### Summary

***DOE's large-scale science projects involve many collaborators at multiple institutions. This leading edge of science depends critically on an infrastructure that supports widely distributed computing and data resources. The DOE Science Grid is being developed and deployed across the DOE complex to provide persistent Grid services to advanced scientific applications and problem solving frameworks. By reducing barriers to the use of remote resources, it is making significant contributions to SciDAC and in deploying the infrastructure required for the next generation of science.***

A significant portion of DOE science is already, or is rapidly becoming, a distributed endeavor involving many collaborators that are frequently multi-institutional. Rapidly increasing data and computing requirements must be addressed with resources that are often more widely distributed than the collaborators. Thus, leading edge science now depends critically on an infrastructure that supports the process of distributed science.

The goal of Grid computing is to deliver:

- Computing capacity adequate for tasks, provided at the time the task is needed by the science;
- Data capacity sufficient for the science task provided independent of location;
- Communication capacity sufficient to support all of the aforementioned and provided transparently to both systems and users, and;
- Software services supporting a rich environment that enables scientists to focus on the science simulation and analysis aspects of software and problem solving systems, rather than on the details of managing the underlying computing, data, and communication resources.

Grid technology is evolving to provide the services and infrastructure needed for building “virtual” systems and organizations. A Grid-based infrastructure provides a way to use and

manage widely distributed computing and data resources in the science environment, and offers an opportunity for a standard, large-scale computing, data, instrument, and collaboration environment for science that spans many different projects, institutions, and countries.

The DOE Science Grid is developing aspects of Grid technology to ensure that it provides a basis of support for DOE's large-scale science projects, with components located at multiple laboratories and universities worldwide. The goal of the DOE SG project is threefold:

- Reduce the barriers by defining common practices for DOE
- Assist in the building of needed infrastructure across sites and projects
- Deploy a testbed initially among five project sites

The project is integrating activities in deployment, research and development, and application outreach that allow us to develop and refine the Grid tools and their deployment and support. DOE SG is focusing on identifying and resolving scalability issues so that the Grid can support large-scale science collaborations. Close cooperation with a variety of application projects is ensuring relevance to SciDAC goals and enabling innovative approaches to scientific computing

via secure remote access to online facilities, distance collaboration, shared petabyte datasets, and large-scale distributed computation. Current application interactions include the Genomes to Life project, the DOE Climate Change Prediction program, and the PLANCK-Grid astrophysics project.

Major accomplishments to date include:

- Construction of a Grid across five major DOE facilities with an initial complement of computing and data resources;
- Integration of NERSC's production, large-scale storage systems into the Grid.
- Creation of the *DOEGrids* PKI [1], a common PKI service for Grid authentication that combines several DOE science VOs, and has been expanded to include several organizations funded by NSF.
- The development of authentication and authorization services, the core to the construction of basic Grid services, called the *International Grid Federation* (IGF) [2]. IGF addresses the direct need for U.S., European, and Asian VOs to work together, and establishes an infrastructure to facilitate the creation and sharing of international Grids, starting with the development of common criteria for identity establishment. The current membership in IGF consists of 20 countries.
- Firewall testing to facilitate deployment of Grid services in firewalled environments. Joint work between NERSC and the Globus Alliance examined various general firewall configurations and resulted in the development of a document explaining how Globus Toolkit® services interact with firewalls[3]. In addition, NERSC worked with the Globus Alliance to make NERSC's intrusion detection system, Bro [4], Globus aware.
- Joint development of a trouble ticket interchange format with the NSF-funded iVDGL. Deployment of a top-level support page, to serve as a single point of contact for submitting trouble tickets. This top-level tool then uses the interchange format to integrate with the local site trouble ticket systems to allow routing to the appropriate site for resolution.

- Design of a resource monitoring and debugging infrastructure that facilitates managing this widely distributed system and the building of high-performance distributed science applications;
- Establishment of development and deployment partnerships with several key vendors, e.g., IBM.
- Development of Python-wrapped Globus services toolkit [5] that enables developers to build experimental Grid system-administration tools to help support the production usage of Grid resources across the DOE SG sites. These include graphical tools for checking the status of Grid resources, managing Globus configuration, adding new Grid users, etc.
- Use of the Grid infrastructure by applications from several disciplines – computational chemistry, ground water transport, climate modeling, bioinformatics, etc.
- A number of demonstrations of DOE Science Grid technology were given at SC 2003

These are important steps in developing and deploying a realistic-scale Grid environment that supports advanced Grid services for DOE science.

For more information visit [www.doesciencegrid.org](http://www.doesciencegrid.org) or contact [kramer@nersc.gov](mailto:kramer@nersc.gov).

#### References:

- [1] DOE Grids PKI, <http://www.doegrids.org/>
- [2] International Grid Federation, <http://www.gridpma.org>
- [3] V. Welch, "Globus Toolkit Firewall Requirements", <http://www.globus.org/security/v2.0/Globus%20Firewall%20Requirements-5.pdf>
- [4] V. Paxton, "Bro: A System for Detecting Network Intruders in Real Time, Proceedings of the 7<sup>th</sup> USENIX Security Symposium, 1998.
- [5] [www-itg.lbl.gov/Grid/projects/pyGlobus/](http://www-itg.lbl.gov/Grid/projects/pyGlobus/)