Enabling Large-Scale Science on the DOE Science Grid

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, the Grid is making significant contributions to SciDAC and in deploying the infrastructure required for the next generation of science.

For the past 30 months, the DOE Science Grid has been developing Grid technology to support DOE's large-scale science projects, with resources located at multiple laboratories and universities worldwide. The Grid's participants are Lawrence Berkeley National Laboratory (LBNL), Pacific Northwest National Laboratory (PNNL), Argonne National Laboratory (ANL), Oak Ridge National Laboratory (ORNL), and the National Energy Research Scientific Computing Center (NERSC), located at LBNL.

Major accomplishments to date (detailed below) include the following:

- Design and deployment of a Grid security infrastructure that is facilitating collaboration between U.S. and European high energy physics projects and within the U.S. magnetic fusion community
- Testing and evaluation of Grid Services on production resources at PNNL, NERSC, and ORNL
- Work on intrusion detection and Grid software's interactions with firewalls
- Design of a Grid resource usage data management and accounting system
- Development and deployment of a cross-site trouble ticketing system

A collaboration between Lawrence Berkeley National Laboratory, Pacific Northwest National Laboratory, Argonne National Laboratory, Oak Ridge National Laboratory, and the National Energy Research Scientific Computing Center

Major Accomplishments

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 public key infrastructure (PKI), a common PKI service for Grid authentication that combined several DOE science virtual organizations (VOs), and has been expanded to include several organizations funded by NSF.

- Development of authentication and authorization services, the core to the construction of basic Grid services, called the International Grid Federation (IGF). IGF addressed 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. In addition, NERSC worked with the Globus Alliance to make NERSC's intrusion detection system, Bro, Globus aware.
- Jointly developed a trouble ticket interchange format with the NSF funded iVDGL. Deployed 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.
- Designed a resource monitoring and debugging infrastructure that facilitates managing this widely distributed system and the building of high performance distributed-science applications.
- Established development and deployment partnerships with several key vendors (e.g., IBM).
- Development of Python-wrapped Globus services toolkit that enables developers to build experimental Grid system-administration tools to help support the production usage of Grid resources across the DOE Science Grid 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, groundwater transport, climate modeling, bioinformatics, etc.
- There were a number of demonstrations of DOE Science Grid technology at SC 2003.

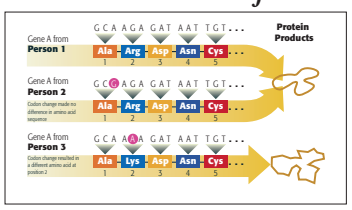
The goal of the next phase of the DOE Science Grid is to develop a

Persistent Science Grid

that will present a consistent, easy to use and stable software environment throughout the participating sites and to support the use of this Grid by key DOE science applications.

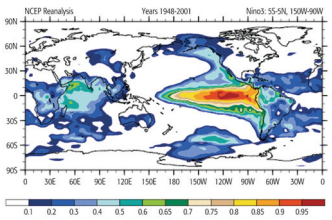
Large Scale Science applications scheduled to be deployed on the Persistent Science Grid

Genomes to Life



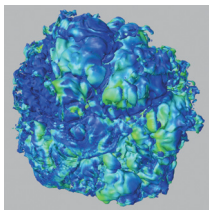
DNA sequence variation in a gene can change the protein produced

Climate Modeling



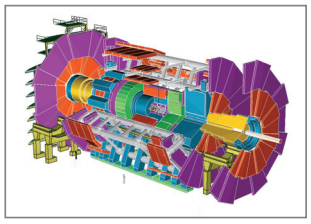
Fifty years of El Niño climate data used to compare against the Community Climate System Model (CCSM2).

Astrophysics



This 3-D supernova simulationshows the turbulent environment beneath the supernova shock wave

High-Energy and Nuclear Physics



Two high-energy experiments that plan to significantly expand their scope of Grid computing over the next few years are ATLAS and ALICE.

Ernest Orlando Lawrence Berkeley National Laboratory · NERSC National Energy Research Scientific Computing Center · ornl Oak Ridge National Laboratory · Argonne National Laboratory University of Chicago · William R. Wiley EMSL · Office of Science U.S. Department of Energy