Middleware Technology to Support Science Portals: a Gateway to the Grid.

Dennis Gannon, PI. Department of Computer Science, Indiana University
Randall Bramley, Co-PI, Department of Computer Science, Indiana University

Summary

A Science Portal is problem solving environment that allows scientists the ability to program, access and execute distributed "Grid" applications from a conventional Web Browser and other desktop tools. In such a portal, scientific domain knowledge and tools are presented to the user in terms of the application science and not in terms of complex distributed computing protocols. The goal is to allow the scientist to focus completely on the science by making the Grid a transparent extension of the user's desktop computing environment.

1. Introduction

Research on Grid portals began in the late 1990s with early attempts to put web-based front-ends on Grid applications and management interfaces. A research group within the Global Grid Forum devoted to Grid Computing Environments began to collect information and requirements. At the same time the concept of web portals and emerged in industry. A portal is a web server that provides each user with a private, configurable environment to manage information resources and tools.

In our project, a Grid portal is a portal that provides interfaces to Grid services and applications conforming to several key standards including the Open Grid Service Infrastructure (OGSI) and its follow-on WSRF, the SciDAC Common Component Architecture (CCA) and the Apache Jetspeed portal model.

Our goal is to provide the user with an "information space", in which he or she may store, search for and discover resources associated with their science and a place where Grid application workflows may be configured launched and controlled. For example, a scientist may need to run an application that consists of several hundred data analysis and simulation tasks. The coordination of this work is through a workflow system. The results of these analysis and simulation tasks are data object that are described in terms of metadata stored in metadata directories. The scientist can later return to the analysis and search through the metadata and study or visualize the results. In this project, we do not define the metadata catalogs or the workflow systems. Many of these exist. Our task is to provide an architecture where it is possible to integrate these tools into a common portal framework.

2. The Collaborations

The research in this project requires two supporting activities in order to be productive. First we need a software platform for building the portal that can be used as a testbed for our new ideas. Second, we need application clients who can give us feedback on how well the research results work when placed in the users hands. For applications, we partnered with the NSF Teragrid, the NSF ITR project on sever storm modeling (http://lead.ou.edu), and application arising from the SciDAC CCTTSS (http://www.cca-forum.org) and the NASA IPG project. Our collaboration

with the Global Grid Forum GCE is now funded by the NSF Middleware Initiative (NMI) to refine the research portal technology developed here for production use. This partnership now also includes the University of Michigan, the University of Texas, Argonne National Laboratory and NCSA. We are building a common infrastructure for Science Portals.

3. The Technology

The current portal prototype provides users with the following set of portal components called portlets.

1. **A Grid Proxy Credentials Manager** that manages the user's authentication certificates so that they be used by the other portlets on the users behalf.
2. **Resource Browsers** that provide the user with the current state of the resources that he or she is authorized to use.
3. **Information Service Interfaces** that provide the user with current information about the user's information space. We refer to this as the MyGrid Context service (see Figure 1).
4. **Job Status Tools** that allow the user the ability to monitor remote jobs.
5. **Messaging Systems,** which allow groups of users to post and read messages about common projects.
6. **Event and logging Systems** that keep track of application events and log them for later discovery from the portal. This also includes tools for journaling and recording event histories as part of experiment metadata called the Grid Context described below.
7. **Portlets to Access Application Factory Services,** which manage the details of complex, multi-component remote applications and workflows.

Our future efforts are focused on building better interfaces to building and launching grid workflows and a more complete set of tools to do information discovery and analysis. Our goal is to make it very easy for groups of collaborating scientists to create shared information spaces and applications in a reliable and secure manner.

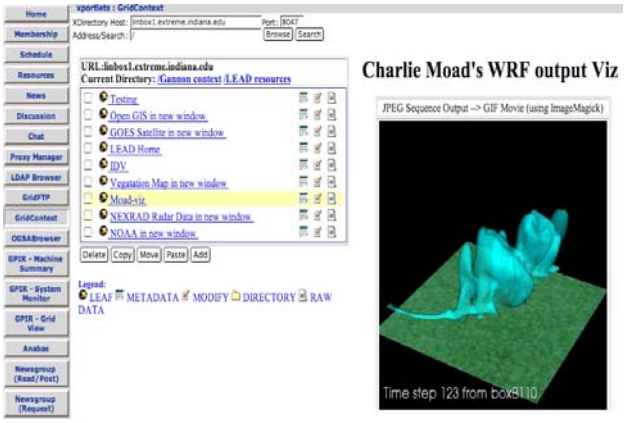

Figure 1. MyGridContext Directory Service. The service stores metadata about grid objects and services. For example, a directory entry can contain metadata about a computational experiment and a HTML link to the visualization resulting from this experiment. In this screen shot, the GridContext entry that is displayed in "Moad-viz" links to a GIF movie of a WRF simulation of a severe storm.

A prototype implementation is available to interested users. It is up and running at http://lead.extreme.indiana.edu:10081/lead/. The software is being distributed and it is available at http://www.ogce.org .

For further information on this subject contact:

Professor Dennis Gannon, Chair
Dept. of Computer Science
Indiana University
Bloomington, IN 47405

Phone:812 855 4510
E-mail: gannon@cs.indiana.edu