



Building a Portal Middleware Grid for Geoscience Communities

Sandeep Chandra, Chaitan Baru (SDSC), Arun Agarwal, K. V. Subbarao (UofHyd)
San Diego Supercomputer Center, UCSD (chandras@sdsc.edu)

A fundamental objective of the GEON Project (www.geongrid.org) is to develop data sharing frameworks—in the process, identifying best practices and developing capabilities and tools to enable advances in how geoscience research is done. The GEON portal framework, which plays a key role in facilitating this objective, is implemented using a portal based framework that provides a uniform authentication and authorization environment with access to a rich set of functionality, including the ability to register data and ontologies into the system; smart search capability; access to a variety of geoscience-related tools; access to Grid-enabled geoscience applications; and a customizable private work space from which users have access to the scientific workflow environment, which allows them to easily author and execute processing “pipelines” for data analysis, modeling, and visualization. In this paper, we describe the modular hardware and software components that we have developed, which make it possible to easily deploy a data sharing portal middleware environment for any application domain. We then describe how this framework was successfully deployed to build a geosciences grid, called iGEON-India, in India.

In practice, deploying a portal framework like GEON portal requires an understanding of the various middleware components, and their dependencies, that go into engineering such a system. In addition to sites within the GEON network, the GEON software infrastructure is increasingly being adopted in other projects such as the Chesapeake Bay Environmental Observatory (CBEO), the Network for Earthquake Engineering Simulations (NEES), the Digital Archeology project (TDAR), and the National Ecological Observatory Network (NEON). Based on this experience, we have developed a modular packaging of the various components of the system to allow easy installation

and configuration of the hardware, middleware, and other software components.

The data portal middleware infrastructure consists of the following components: 1. A Portal Server, which runs the portal software, The nominal system is a “rack-mounted”, server-class machine with 750 GB raw disk (5 x 146 GB hot-swappable SAS drives), dual core 3.0 GHz Intel Xeon processors, 4 GB of RAM, dual gigabit network interfaces, and redundant power supplies. The portal server runs the GEON software stack, including the portal software and provides connectivity to and interoperability among the other GEON systems.

2. A Data Server, which provides storage and other data management services. This system is also a “rack-mounted”, server-class machine with dual core 3.0 GHz Intel Xeon processors. It includes 5 x 300 GB RAID SAS drives for a total of 1.5 Terabytes of raw drive space. The data nodes are configured with RAID disks in order to deal with unforeseen disk failures.

3. A Certificate Authority (CA) Server, which manages user accounts. The CA server system has the same basic configuration as the Portal Server, but with 2 GB RAM and the disk size reduced to about 36 GB, since the CA Server tasks are not I/O intensive.

The Portal Server runs the Rocks cluster management software and a standardized “GEON software stack,” which includes the GEON Portal and its dependent libraries, including software tools developed in the GEON project. The so-called “core” portal functionality is generic, e.g. search and data ingestion capabilities, and can mostly be leveraged “out of the box” by other projects. The Data Server provides the capability to host data registered through the portal and also provides other data management services. The Data Server runs the SDSC Storage Resource Broker (SRB) software, which provides a number of built-in data management services. The Data Server could also host additional data management services, e.g. GIS software. The Portal Server can communicate with Web services hosted at remote locations using the standard web service protocols.

The CA Server runs the Grid Account Management Architecture (GAMA) software, which manages users accounts through the portal. The CA server is installed using the Rocks software and the GAMA roll. Once installed, the system is fully configured as a CA. Also, the portal software is pre-configured to communicate with this GAMA server for managing user accounts.

The goal of iGEON-India is to promote the use of cyberinfrastructure at participating institutions in India to facilitate the sharing of geosciences data and tools via GEON middleware. The lead institution in India, which is the University of Hyderabad, has deployed the GEON portal middleware at their site in Hyderabad to enable local scien-

tists to register their data sets. Consequently, similar GEON resources were deployed at two other locations namely, University of Pune and Adikavi Nannaya University at Rajahmundry.

Deploying the GEON portal middleware framework at these distributed sites provides capabilities to (i) develop and deploy services, tools and serve datasets for the users at their site and for the broader iGEON-India community, (ii) help the partner site integrate local department or campus resources, and (iii) provide a platform for other research organizations and universities to come together and collaborate. Users at these universities have already started registering some data sets, including GIS data, geo-chemistry data and various other applications and tools.