Towards Service-Oriented Geoscience: SEE Grid and APAC Grid

R. Woodcock[(1)], R. Fraser[(2)]

(1) CSIRO Exploration and Mining and pmd*CRC, 26 Dick Perry Ave, Kensington, Western Australia 6151, Australia
(2) IVEC, 26 Dick Perry Ave, Kensington, Western Australia 6151, Australia

Abstract

Open geospatial standards, service-oriented architectures (SOA) and grid computing enable new approaches to publishing and accessing geoscience data and programs. The linkages between three collaborating projects show how the use of standards based service interfaces and protocols is enhancing the capabilities of geoscientists. The first of these projects, "The Solid Earth and Environment Grid (SEE Grid) Roadshow 2005" demonstrates how open geospatial standards can be used to provide interoperable access to Government geoscience data held at all Australian geological surveys. The "Australian Partnership for Advance Computing (APAC) Grid Geosciences" project illustrates how computationally demanding geoscience programs can be made available as services and distributed across the APAC partners HPC and storage resources in a manner that requires limited knowledge of the physical infrastructure. Finally, the "predictive minerals discovery CRC (pmd*CRC) project" shows how these services can be chained to provide advanced modelling and interactive inversion of mineralization processes for the purposes of improved exploration targeting.

Keywords: Service-oriented architecture; grid computing; computational geoscience

Introduction

CSIRO Exploration and Mining and the predictive minerals discovery CRC (pmd*CRC) use numerical modelling of geological systems to systematically explore the process-related parameters governing the formation of mineral deposits (http://www.pmdcrc.com.au/). At a high level, the workflow used is common with many other research investigations:

1) gather input data including geological and geometric properties and create the model
2) perform the computation using a suitable program and computing architecture
3) analyse the results
4) repeat if further study is required

The types of investigations undertaken are highly variable ranging from simplified geologic models to real-world scenarios. Input data includes geological observations of rock properties and chemical composition supplied by mining companies and geological surveys. Finite element meshes are created from 3D geological models often supplied by mining companies and can take 1 – 2 weeks to produce. Computation time may range from a couple of hours on a desktop PC to a couple of weeks on an HPC depending on which phenomena are simulated, the numerical solver, the computing architecture, and the number of studies required to explore the parameter space. The result is a demanding workflow with flexibility and efficiency being required at *all* stages.

In order for the approach to be cost effective investigations are often simplified at various stages. For example, the use of interactive inversion [Boschetti 2001] can greatly improve the investigation process by "optimising" the parameters towards models that are "better" under the guidance of the geoscientist. This approach effectively eliminates large sections of the parameter space and substantially reduces computational cost. The use of template-based numerical modelling [Potma 2004] can be used to automatically create families of geometrically related models based on a single manually created model. This automation of the workflow significantly reduces the time required to create the models and improves overall productivity.

However there are significant inefficiencies in the workflow that have not be eliminated. These inefficiencies occur with interactions between people, organisations and resources:

- Information scattered across multiple geological surveys and the mining companies hampers the gathering of input data. Consequently, the cost of data integration can substantially exceed all other costs. Investigators often ignore this wealth of real world observation data preferring to use "average" properties that could range by several orders of magnitude between geographic locations.
- Multiple HPC resources are available, particularly for research purposes, via the Australian Partnership for Advance Computing (APAC). However, access is often difficult due to differing queuing and data staging policies at each site. Investigators often find the cost of adapting their tools to use multiple sites prohibitive and limit their access

to resources either to their own PC's or a single HPC facility.

We are involved in several projects that, when linked, substantially address the issues of scattered information and access to multiple HPC resources in the geosciences. Further we apply these technologies to the pmd*CRC modelling of mineralization processes for the purposes of improved exploration targeting.

These issues are seen across multiple domains, not just the geosciences. As a result, open geospatial standards (http://www.opengeospatial.org), service-oriented architectures (http://www.ibm.com/developerworks/webservices) and grid computing [Foster 2001] have been developed as domain-independent technologies to assist in dealing with these issues.

First we will describe how open geospatial standards have been used in the Solid Earth and Environment Grid to provide access to pre-competitive geoscience data. Then we will describe our activities in the development of the APAC Grid. Finally, we will illustrate how the two previous projects are used in pmd*CRC software architecture.

The Solid Earth & Environment Grid Roadshow 2005

The Solid Earth & Environment Grid (SEE Grid) Roadshow project's aim was to provide on demand, web service based access to geoscience information holdings at State and Territory Geological surveys using a *common* service interface and information model. This was to be done without alteration to the backend database technologies and private schemas of the surveys. Participants in the project included the CSIRO, Geoscience Australia, Social Change Online, all State and Territory Geological Surveys with support from AusIndustry and the Minerals Council of Australia.

The implementation made use of the Open Geospatial Consortium Web Feature Service (WFS) for the common service interface [Vretanos 2004] and the CSIRO's Exploration and Mining Mark-up Language (XMML) for the information model (https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome). The WFS middleware was implemented as an extension to the open source Geoserver WFS (http://geoserver.sourceforge.net/html/index.php). The system was deployed at each state and territory survey with only minor modifications being required to configure the mapping from private data sources (restricted to geochemical assay data for this project) to the community information model. Three client applications, a web based GIS, web based text report, and a commercial mining software package, were also modified to use the WFS services.

At the conclusion of the project all 8 geological surveys had successfully implemented and deployed the WFS service. The client applications could then trivially interrogate all the surveys for geochemical assay information regardless of state and territory boundaries and local information models.

This open standards approach was demonstrated in all state and territory capital cities to industry and government stakeholders. The overwhelming response was that the demonstration clearly showed that the issue of data integration and access can be substantially solved through the SEE Grid approach. The CSIRO and Geological Surveys are now developing production grade services for other geological data types via the SEE Grid community of practice. The full availability and improved access to this geoscience data will greatly benefit both industry and research.

APAC Grid: Geosciences

The Australian Partnership for Advanced Computing, which has partners in most state and territories, provide HPC and mass storage facilities for research purposes. These facilities are managed independently and have a range of computing hardware and software systems which require users to adapt their workflow to a specific facility. In order to better facilitate access, APAC has been implementing Grid technology to provided standardised access to domain-independent services like job management, data storage, monitoring, and security. This standardisation goes some way towards improving access to the resources at the facilities regardless of the underlying software and hardware infrastructure.

In addition to this, APAC is supporting the development of discipline-specific services in areas like bioinformatics, astronomy, high energy physics and the geosciences. The APAC Grid Geosciences project is developing service interfaces for commonly used algorithms and making them available at multiple locations using the standard interface. The geoscience specific services are built on top of the domain-independent grid technology and, as a result, development of the geoscience specific services is greatly accelerated. Researchers can make use of these services by chaining them together to support their workflow.

As a demonstration of the approach several applications are being deveoped. A seismic simulator web portal and computational service has recently been completed. A mantle convection service based on the Snark numerical code is under development. Ultimately the goal is to service chain the mantle convection service with an independently developed interactive inversion service. Finally, the EarthByte 4D data portal will provide a service for simulation and visualisation of geological and geophysical observations coded by tectonic plate and geological time. The EarthByte data portal will retrieve the observational data from the geological surveys using the SEE Grid approach. In addition, the results of the EarthByte simulation can be chained into the mantle convection service to provide initial and boundary conditions for modelling runs. This service chaining illustrates how the use of service based architectures can support the creation of more complex systems with substantially reduced development effort. The project is also developing a "how-to" guide to assist other groups who wish to make their programs available as grid enabled services.

The standardisation of interfaces across the APAC Grid will allow a greater number of researchers to access more resources. In addition, the publication of discipline-specific services that can be chained together will facilitate better access to research outcomes and further collaboration.

pmd*CRC Modelling Toolkit

The pmd*CRC modelling workflow was briefly described in the introduction to this article. It is a demanding workflow, often requiring integration of disparate data, substantial computation and the flexibility to substitute numerical algorithms better suited to the investigation at hand. The developments being undertaken in the SEE and APAC Grid enable the cost effective development of a modelling toolkit to support this workflow. An example of possible service interactions is shown in Figure 1. This would be a challenging and costly development if all components needed to be developed and deployed by a single organisation. However, the SEE and APAC Grid communities are developing many of these components already. The main focus of the pmd*CRC development is on the workflow and the orchestration of the service interactions.

Figure 1: *An example of the service interactions involved in the pmd*CRC modelling toolkit. With SEE Grid and APAC Grid fully deployed, only the workflow component requires development, all other services pre-exist. Services can be substituted e.g. Snark replaced with a more suitable numerical code for a given problem.*

Discussion

The formation of communities of practice, like the SEE and APAC Grid, who publish useful services using open standards, provides new opportunities for the publication and access to geoscience data and programs. We have shown that such services can be developed, deployed and orchestrated in order to support advanced geoscience modelling workflows.

References

F. Boschetti and L. Moresi, Interactive Inversion in Geosciences, *Geophysics*, 64, 1226-1235, 2001

I. Foster, C. Kesselman and S. Tuecke, The Anatomy of the Grid, *Intl. J. Supercomputer Applications*, 2001

W. Potma, P. Schaubs and T. Poulet, Application of template-based generic numerical modelling to exploration at all scale, *pmd*CRC Conference*, Barossa Valley, 2004

P. Vretanos, Web Feature Service 1.1, OGC document 04-094, 2004