



Geoscientific Workflows using Grid Computing Infrastructure

Ryan Fraser, Robert Woodcock, and Terry Rankine
CSIRO Earth Science and Resource Engineering

A major benefit of high-performance computing is the vast improvement it offers scientists and researchers in Australia when it comes to exploring large data problems. Thus, high performance computers and grid computing have become key components in exploration and mining applications in the past few years; anything from running numerical simulations to creating a virtual laboratories or providing a mechanism to discover and process geodetic information collected by GPS stations.

However, scientists have previously been restricted in their abilities to access high-performance computing for these and other uses due to the fragmented nature of the infrastructure. Australia's high-performance centres host a range of different supercomputers from different manufacturers, with differing configurations, different architectures. Each supercomputer is different so users who want to use more than one supercomputer successfully need to know the difference between them all.

CSIRO has found an answer to this dilemma via the AuScope Grid. AuScope Grid is creating an e-Research Infrastructure to federate and make nationally distributed datasets and high-performance computing resources interoperable. AuScope Grid is developing tools to manipulate large data volumes and establishing an appropriate governance framework to ensure sustainability.

AuScope Grid's premise is to comprise distributed data storage hardware, high bandwidth network links, data management protocols, middleware and software. Major geoscience and geospatial data stores of the government agencies are deploying this technology for use internally and as an external face to their data. Combining this with high-performance compute resources and high-bandwidth networks the academic community can now tackle some of the science problems they have wanted to attempt for some time but have not been feasible until now.

AuScope has deployed a grid computing platform which standardises access to high performance computers regardless of the machines' type, manufacturer and location.

Grid computing is "middleware infrastructure", the essential catalyst that sits between a high-performance computer and a workflow client, such as the Virtual Rock Laboratory (a virtual space to crush rocks or particles) that simplifies access to high performance compute resources.

Workflows were possible before the deployment of grid computing infrastructure but some required an order of magnitude longer to produce the desired outcome (differences of up to 6 months have been recorded) and typically required the user to have a degree in Computer Science. The Virtual Rock Laboratory, Geodesy Workflow and the Desktop Modelling Toolkit are all examples of workflow clients developed to enable complex scientific workflows. Grid computing has aided the development of these workflows and made it possible for scientists to process large scientific problems in record times.