Geophysical Research Abstracts
Vol. 14, EGU2012-7060, 2012
EGU General Assembly 2012

Working Towards New Transformative Geoscience Analytics Enabled by Petascale Computing

R. Woodcock (1) and L. Wyborn (2)
(1) CSIRO, Canberra, Australia, (2) Geoscience Australia, Canberra, Australia (lesley.wyborn@ga.gov.au)

Currently the top 10 supercomputers in the world are petascale and already exascale computers are being planned. Cloud computing facilities are becoming mainstream either as private or commercial investments. These computational developments will provide abundant opportunities for the earth science community to tackle the data deluge which has resulted from new instrumentation enabling data to be gathered at a greater rate and at higher resolution. Combined, the new computational environments should enable the earth sciences to be transformed.

However, experience in Australia and elsewhere has shown that it is not easy to scale existing earth science methods, software and analytics to take advantage of the increased computational capacity that is now available.

It is not simply a matter of 'transferring' current work practices to the new facilities: they have to be extensively 'transformed'. In particular new Geoscientific methods will need to be developed using advanced data mining, assimilation, machine learning and integration algorithms. Software will have to be capable of operating in highly parallelised environments, and will also need to be able to scale as the compute systems grow. Data access will have to improve and the earth science community needs to move from the file discovery, display and then locally download paradigm to self describing data cubes and data arrays that are available as online resources from either major data repositories or in the cloud.

In the new transformed world, rather than analysing satellite data scene by scene, sensor agnostic data cubes of calibrated earth observation data will enable researchers to move across data from multiple sensors at varying spatial data resolutions. In using geophysics to characterise basement and cover, rather than analysing individual gridded airborne geophysical data sets, and then combining the results, petascale computing will enable analysis of multiple data types, collected at varying resolutions with integration and validation across data type boundaries. Increased capacity of storage and compute will mean that uncertainty and reliability of individual observations will consistently be taken into account and propagated throughout the processing chain.

If these data access difficulties can be overcome, the increased compute capacity will also mean that larger scale, more complex models can be run at higher resolution and instead of single pass modelling runs. Ensembles of models will be able to be run to simultaneously test multiple hypotheses.

Petascale computing and high performance data offer more than "bigger, faster": it is an opportunity for a transformative change in the way in which geoscience research is routinely conducted.