Geophysical Research Abstracts
Vol. 14, EGU2012-4491, 2012
EGU General Assembly 2012
© Author(s) 2012

The FOSS GIS Workbench on the GFZ Load Sharing Facility compute cluster

P. Löwe, J. Klump, and J. Thaler

Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, Centre for Geoinformation Technologies, Potsdam, Germany (peter.loewe@gmx.de)

Compute clusters can be used as GIS workbenches, their wealth of resources allow us to take on geocomputation tasks which exceed the limitations of smaller systems.
To harness these capabilities requires a Geographic Information System (GIS), able to utilize the available cluster configuration/architecture and a sufficient degree of user friendliness to allow for wide application.
In this paper we report on the first successful porting of GRASS GIS, the oldest and largest Free Open Source (FOSS) GIS project, onto a compute cluster using Platform Computing's Load Sharing Facility (LSF).
In 2008, GRASS6.3 was installed on the GFZ compute cluster, which at that time comprised 32 nodes. The interaction with the GIS was limited to the command line interface, which required further development to encapsulate the GRASS GIS business layer to facilitate its use by users not familiar with GRASS GIS.
During the summer of 2011, multiple versions of GRASS GIS (v 6.4, 6.5 and 7.0) were installed on the upgraded GFZ compute cluster, now consisting of 234 nodes with 480 CPUs providing 3084 cores. The GFZ compute cluster currently offers 19 different processing queues with varying hardware capabilities and priorities, allowing for fine-grained scheduling and load balancing.
After successful testing of core GIS functionalities, including the graphical user interface, mechanisms were developed to deploy scripted geocomputation tasks onto dedicated processing queues. The mechanisms are based on earlier work by NETELER et al. (2008).
A first application of the new GIS functionality was the generation of maps of simulated tsunamis in the Mediterranean Sea for the Tsunami Atlas of the FP-7 TRIDEC Project (www.tridec-online.eu). For this, up to 500 processing nodes were used in parallel.
Further trials included the processing of geometrically complex problems, requiring significant amounts of processing time. The GIS cluster successfully completed all these tasks, with processing times lasting up to full 20 CPU days.
The deployment of GRASS GIS on a compute cluster allows our users to tackle GIS tasks previously out of reach of single workstations. In addition, this GRASS GIS cluster implementation will be made available to other users at GFZ in the course of 2012. It will thus become a research utility in the sense of "Software as a Service" (SaaS) and can be seen as our first step towards building a GFZ corporate cloud service.