



Combining the GRID with Cloud for Earth Science Computing

Dmitry Mishin, Oleg Levchenko, Andrei Groudnev, and Mikhail Zhizhin

Geophysical Center Russian Academy of Sciences

Cloud computing is a new economic model of using large cluster computing resources which were earlier managed by GRID. Reusing existing GRID infrastructure gives an opportunity to combine the Cloud and GRID technologies on the same hardware and to provide GRID users with functionality for running high performance computing tasks inside virtual machines. In this case Cloud works “above” GRID, sharing computing power and utilizing unused processor time. We manage virtual machines with Eucalyptus elastic cloud and we use Torque system from gLite infrastructure for spreading Cloud jobs in GRID computing nodes to scale the parallel computing tasks on virtual machines created by elastic cloud. For this purpose we have added new types of tasks to the standard GRID task list: to run a virtual node and to run a job on a virtual node. This gives a possibility to seamlessly upscale the Cloud with the new tasks when needed and to shrink it when the tasks are completed. Using GRID components for managing the size of a virtual cloud simplifies building the billing system to charge the Cloud users for the processor time, disk space and outer traffic consumed.

A list of Earth Science computing problems that can be solved by using the elastic Cloud include repetitive tasks of downloading, converting and storing in a database of large arrays of data (e.g. weather forecast); creating a pyramid of lower resolution images from a very large one for fast distributed browsing; processing and analyzing the large distributed amounts of data by running Earth Science numerical models.