Geophysical Research Abstracts,
Vol. 10, EGU2008-A-11319, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-11319
EGU General Assembly 2008

GRID workflow management for climate simulations.

V. Fernandez-Quiruelas, J. Fernández, A. S. Cofiño and J. M. Gutiérrez, Rafa Mayo
Dep. of Applied Mathematics and Computer Sciences, University of Cantabria, Spain

Climate models are a challenging computing problem, since they usually require large amounts of CPU (complicated equations solved), memory (large multi-dimensional arrays) and disk space (nearly all data produced needs to be stored). Moreover, the CPU usage is not only intensive (this is shared with many other GRID applications) but also sustained for periods which could reach several months.

Nowadays it's not common the use of GRID Computing to run long term jobs due to the high rate of failure in jobs sent to GRID and the CPU-time limitations for the jobs on the local management system. This rate of failure is not really important when the task run in the GRID are short, but becomes critical in longer tasks where after some hours calculating we lose our work.

Unlike many other applications ported to GRID, these climate models need to make use of advanced techniques in workflow management.

In this work we provide the following capabilities to manage the simulations. Failure Aware: Due to the nature of GRID there are several reasons why our josbs can fail: geographical distribution of middleware services, heterogeneity, CPU-Time limited queues, misconfigurations.

Restart and checkpointing from failed jobs. All simmulations are producing intermediate results, then our managament system implements a restart system.

Monitoring: Since climate model simulations last for a long time, we need to know what is happening with the simulation once it has been sent to the testbed. Is the model is running?, which step of time is being calculated?, which files were uploaded to Storage Elements?, which is the last restart point?, ...

Data and Metadata storage: The management system also is controlling the generation of output information that can be easily accessed for our users.

These capabilities are based on gLite middleware (AMGA, LFC, see http://glite.web.cern.ch/glite/ for more information) and other tools like GridWay metascheduler (http://www.gridway.org), and the the system has been used in different testbeds like EELA (http://www.eu-eela.org) and EGEE (http://www.eu-egee.org/)

The requirements of the climate models satisfied in this work are only an example of the needs of many other Earth Science applications.