Support for Astrophysical applications in Slovakia

Ladislav Hluchy, Jan Astalos

Institute of Informatics

Slovak Academy of Sciences

www.eu-egee.org

Information Society and Media

Overview

- Support for Astrophysical application in EGEE-II
- Framework for parameter studies
- Planned applications for EGEE-III
- Typical scenario
- Expertise from other Grid projects that can be used in AA cluster

Support for AAA in EGEE II

- **The unified theory of Kuiper-belt and Oort-cloud formation**
  - Neslušan L., Jakubík M., Paulech, T. - AI SAS – EGEE
  - Leto, G. – Catania Observatory – TriGRID
- **Main goal**
  - working out a unified theory of the formation of all:
    - Jovian planets (Jupiter, Saturn, Uranus, and Neptune)
    - Kuiper belt and Scattered Disc (populations of small bodies beyond the Neptune's orbit)
    - Oort cloud (very distant cometary reservoir)
- **Method**
    - simulation of the dynamical evolution of a large number (~10000) planetesimals (treated as test particles) in the proto-planetary disc; the trajectories of the particles are influenced by the perturbing forces from the giant planets, Galactic tide, and stars passing near or through the Oort cloud

More about the application

- **Demands for CPU time**
  - If a single 2.8GHz CPU was used, the computation of the orbits of 4 giant planets and 10038 test particles for 1 Gyr would last about 21 years
  - After the division of the computation to 240 tasks, each for a single CPU, the above mentioned computation can be performed within ~ 5 months
- **Structure**
  - Sequence of sub-simulations
  - Each sub-simulation consists of many independent tasks (180 EGEE/VOCE + 60 TriGRID)
  - Output from all tasks is needed for preparation of next sub-simulation

Parameter studies

- **Main problem - reliability**
  - Failed jobs need to be resubmited – decreasing throughput and productivity
  - Too much effort needed for analysing errors in Grid middleware or sites
  - Need to identify failed and waiting jobs ASAP.
- **Solution used – pilot jobs**
  - Generic, easy-to-use framework for pilot jobs was developed
  - Pilot jobs are running simulation model in cycle with input datasets obtained from working area in Storage Element
  - Monitoring information is being uploaded periodically to Storage Element – heartbeat + progress monitoring
  - Automatic job management - failed/waiting pilot jobs are cancelled and sites are added to blacklist, no user intervention needed.

Use case

1) User prepares input datasets on User Interface and application executing script.
2) User prepares working directory on Storage Element and transfers input datasets into it.
3) User prepares configuration files for job management script.
4) User starts job management script on UI.
5) User occasionally checks status of processing by obtaining monitoring information from SE.
6) Application runs until processing of all input datasets is finished.
7) User collects output datasets from SE to UI and based on them he prepares input data and parameters for next sub-simulation
8) User continues from point 2.) until whole sequence of sub-simulations is finished.

To be done

- **Automatic detection of anomalies in processing**
  - If some tasks run for unusually long or short time – it might indicate some problems at worker nodes
  - Periodic uploading of stdout/stderr to Storage Element – RFIO based utilities are ready, they will be integrated to the framework including automatic detection of errors.
- **Notification of user when intervention is needed**
  - E-mail, ICQ, ...

Applications planned for EGEE-III

- **Investigation of the structure of meteoroid streams (M. Jakubík[1], J. Svoreň[1], Z. Kaňuchová[1], V. Porubčan[1,2];**
  - [1]AISAS, [2]Astron.-Geophys. Obs. Commenius Univ.)
- **Transitions between the sub-populations of small bodies in the solar system**
  - (E. Pittich, N. Solovaya; AISAS)
- **Study of the stability of orbits of exo-planets in multiple systems**
  - (M. Zboril; AISAS)

Typical scenario

- **The described future applications includes numerical experiments; a typical procedure consists of following steps:**
  - Creating the models of a variety of initial situations (choosing the initial values of free parameters in each model)
  - Computing the evolution for each model (typically by utilizing a single CPU)
  - Comparison the obtained final state with the well-known facts (observed reality) and evaluation of which model is the best for the approximation of the reality

Technologies we can support

- **II SAS has experience with variety of Grid technologies from participation in Grid projects:**
  - Development and customisation of scientific Grid portals
  - Porting of advanced applications to Grid - workflow, semantic composition, MPI
  - Data management - metadata management, relational and XML database access (OGSA-DAI)
  - Interactive Grid applications
  - Visualisation of Grid applications
- **Projects:**
  - CrossGrid project - http://www.crossgrid.org
  - KWfGrid project - http://www.kwfgrid.eu
  - MediGrid project - http://www.eu-medigrid.org
  - Int.eu.grid – http://www.interactive-grid.eu