egee

Enabling Grids for E-sciencE

Contribution ID: **26**   Type: **Oral contribution**

Requirements of Climate applications on Grid infrastructures; C3-Grid and EGEE

Wednesday, 1 March 2006 17:45 (15 minutes)

Human made climate change and its impact on the natural and socio-economic environment is one of todays most challenging problems of mankind. To understand and project processes, changes and impacts of the natural and socio-economic system a growing community of researchers from various disciplines investigates and analyses the earthsystem by means of computer simulation and analysis models.
These models are usually computational demanding and data intensive as they need to compute and store high resolved 4-dimensional fields of various parameters. Moreover, the required close collaboration in interdisciplinary and often also international research projects involves intensive community interactions.
To support climate workflows the community established proprietary, mostly national or regional solutions, which are normally grouped around centralized high performance computing and storage resources. Homogeneous discovery of and access to climate data sets residing in distributed petabyte climate archives as well as distributed processing and efficient exchange of climate data are the central components of future international climate research. Thus, the EGEE infrastructure potentially offers a highly suitable environment for such applications.

However, existing grid infrastructures - including EGEE - do not yet meet the requirements of the climate community essential for prevalent workflows. Hence, to port existing applications and workflows on the EGEE infrastructure, a stepwise extension of the infrastructure to community specific services is needed. Moreover, the identification and demonstration of feasibility and added value is essential to convince the community to change their established habits. The Collaborative Climate Community Data and Processsing Grid (C3-Grid [1]) is an application driven approach towards the deployment of GRID techniques for climate data analysis. Solutions currently developed in this project offer a potentially fruitful basis to improve the suitability of the EGEE infrastructure as a platform for data analysis within climate research.

Within EGEE climate is part of the Earth Science Research (ESR) VO. We evaluated and tested the use of the EGEE infrastructure for climate applications [4]. As part of this prototypes of simulation as well as analysis software were tested on the EGEE infrastructure. We identified 3 different accesspoints for pilot applications, that can demonstrate the potential benefit of the EGEE infrastructure for climate research: Ensemble simulations with models of intermediate complexity, coupling experiments on a common platform and data sharing and analysis.

Ensembles of simulations performed with the same model but different future scenarios and different parameterisations are required to quantify the uncertainty and possible variety of future climate predictions. EGEE offers a good infrastructure for such ensemble simulations with models of intermediate complexity, which do not need the performance of a supercomputer. Ensembles can be submitted as DAG, parametric or collection job and results could be directly stored, analysed and reduced to the required information on the grid.

The coupling of diverse models of different disciplines is essential to understand the interaction and feedback between the different climate and earth system components, as e.g. the human impact on future climate development. In corresponding projects partners from different institutes of different nations are collaborating on a common modeling framework. The EGEE infrastructure would be a valuable platform for such coupling approaches. Data, models and output could be easily shared, different access and user rights can be established via VOMS. Currently different coupling tools are explored to assess their "grid-suitability".

Data sharing and analysis is a central aspect in climate research. The enormous amounts of data, produced by the model simulations need to be analysed, visualised and validated against observations or other data sources to be correctly interpreted. This involves a multiplicity of statistical calculations carried out on samples of different large data files. Currently such data analysis is centred around the heterogeneous database systems, which are accessed via non-standardised metadata. Thus, the establishment of a common data exchange and management infrastructure bridging the existing heterogeneous community datamanagement solutions with the EGEE data management system would add great value to such applications.

Especially for the realisation of climate data sharing and analysis workflows on EGEE the following components need to be developed:

1) a common agreed upon metadata schema for discovery of climate data sets stored in grid file space as well as in external community datacenters
2) a common community metadata catalogue based on this schema
3) common interfaces to reference and access grid external data resources (mainly databases)

All of these aspects are addressed within the recently introduced national German C3Grid [1] project within the German e-science (D-Grid [2]) initiative which aims to develop a grid middleware specific for the needs of the climate research community. Within this project a common metadata schema is defined. A community metadata catalogue and information system is established and a common data access interface will be defined.

To promote EGEE as a climate data handling (and postprocessing) infrastructure based on these developments we propose a stepwise approach:

- establishment of an international standards based climate metadata catalog (e.g. based on AMGA plus a common push/pull metadata exchange to grid external metadata catalogues via established metadata harvesting protocols
- establishment of data access to (initially free) climate datasets in climate data centers: As intial starting point we need an easy way to access data in climate data centers and copy/register them on grid storage, e.g. by using proprietary access clients or OGSA-DAI.
- adaptation of commonly used climate data processing toolkits on EGEE such as e.g. cdo [3]

[1] http://www.c3grid.de
[2] http://www.d-grid.de
[3] http://www.mpimet.mpg.de/~cdo/
[4] Stephan Kindermann, EGEE infrastructure and Grids for Earth Sciences and Climate Research, Technical report DKRZ (available under http://c3grid.dkrz.de/moin.cgi/PublicDocs)

Primary author: Dr BIERCAMP, Joachim (DKRZ)

Co-authors: Mrs RONNEBERGER, Kerstin (DKRZ); Dr KINDERMANN, Stephan (DKRZ)

Presenter: Dr BIERCAMP, Joachim (DKRZ)

Session Classification: 1c: Earth Observation - Archaeology - Digital Library

Track Classification: Earth Observation - Archaelogy - Digital Library