



Contribution ID: 69

Type: Oral

Carbon Dioxide Flux Data Computing and Data Grid Warehouses Using Grid Techniques

Wednesday 4 March 2009 14:20 (20 minutes)

In studies of climate change, utilizing the flux observation tower is one of the important research methods. In the FLUXNET, there are over 500 tower sites operating on a long-term and continuous basis in the world. Scientists face a challenge of dealing with a huge amount of data. The EGEE infrastructure is helpful for this research to provide a better environment for huge sensor data management and a computation model with flexible user control. A workflow engine was integrated with gLite by the GAP.

Impact

FLUXNET provides information and data exchange between the global flux community and the broader research communities interested in carbon cycle science and global climate change. The increasing atmospheric concentration of carbon dioxide and the relationship to ecosystems is well known. However, the ecological data collected from the sensors of different sites are stored in different databases that does not have the standard data schema. Furthermore, the current storing and processing data sets of the research model is difficult to share and process using the various tools within FLUXNET, although the researchers have the same measurement process. In this prototype, the EGEE infrastructure was proved to be a good solution for carbon flux computation with intensive and distributed data, computing and knowledge.

URL for further information

<http://gap.grid.sinica.edu.tw>

Conclusions and Future Work

At present, The GUI of the carbon flux data warehouse is available. The user can manage the data with workflow management flexibility over the Grid infrastructure. We also made use of R to provide mathematical computation functions for statistical computing, and deployed an R service on gLite, and thus other domain scientists also can use this service. In the future, the flexibility to integrate more tools, the improvement of user interface, and the reusability of workflow components are the focus for advancement.

Keywords

Carbon dioxide flux, eddy covariance, scientific workflow, Grid Application Platform

Detailed analysis

Quality and flexibility are the major issues for porting applications to a service Grid such as EGEE. When working with carbon flux research groups, flexible control to effectively exercise users computing models is essential. Our solution integrates GAP (Grid Application Platform) and Kepler together with gLite to support the carbon dioxide flux computing and data warehousing for the production EGEE e-Infrastructure. The GAP presents high level APIs for core Grid services of the environment. Many core components for this application were developed to support workflow management over gLite, such as access authentication, job monitoring, data discovery, and data access services, etc. Furthermore, we developed the Carbon Flux job submission

workflow with Kepler based on the computing model of the user groups, from data preparation, analysis job submission and monitoring, data management, and archives, etc.

Authors: Mr HSU, Cheng-Hsin (Academia Sinica Grid Computing); Mr UENG, Wei-Long (Academia Sinica Grid Computing)

Co-authors: Dr LIN, Chau-Chin (Taiwan Forestry Research Institute); Mr YEN, Eric (Academia Sinica Grid Computing); Mr CHEN, Hsin-Yen (Academia Sinica Grid Computing); Dr LIN, Simon (Academia Sinica Grid Computing); Dr HSIA, Yue-Joe (Institute of Natural Resources National Dong Hwa University)

Presenter: Mr HSU, Cheng-Hsin (Academia Sinica Grid Computing)

Session Classification: Earth Sciences

Track Classification: Experiences from application porting and deployment