Geophysical Research Abstracts
Vol. 12, EGU2010-6713-1, 2010
EGU General Assembly 2010

Grid-based platform for training in Earth Observation

Dana Petcu (1), Daniela Zaharie (1), Silviu Panica (1), Marc Frincu (1), Marian Neagul (1), Dorian Gorgan (2), and Teodor Stefanut (2)

(1) West University of Timisoara, Computer Science Department, Timisoara, Romania (petcu@info.uvt.ro, +40-256-592-316), (2) Technical University of Cluj Napoca, Computers Department, Cluj Napoca, Romania

GiSHEO platform [1] providing on-demand services for training and high education in Earth Observation is developed, in the frame of an ESA funded project through its PECS programme, to respond to the needs of powerful education resources in remote sensing field. It intends to be a Grid-based platform of which potential for experimentation and extensibility are the key benefits compared with a desktop software solution. Near-real time applications requiring simultaneous multiple short-time-response data-intensive tasks, as in the case of a short time training event, are the ones that are proved to be ideal for this platform.

The platform is based on Globus Toolkit 4 facilities for security and process management, and on the clusters of four academic institutions involved in the project. The authorization uses a VOMS service. The main public services are the followings: the EO processing services (represented through special WSRF-type services); the workflow service exposing a particular workflow engine; the data indexing and discovery service for accessing the data management mechanisms; the processing services, a collection allowing easy access to the processing platform.

The WSRF-type services for basic satellite image processing are reusing free image processing tools, OpenCV and GDAL. New algorithms and workflows were develop to tackle with challenging problems like detecting the underground remains of old fortifications, walls or houses. More details can be found in [2].

Composed services can be specified through workflows and are easy to be deployed. The workflow engine, OSyRIS (Orchestration System using a Rule based Inference Solution), is based on DROOLS, and a new rule-based workflow language, SILK (SImple Language for worKflow), has been built. Workflow creation in SILK can be done with or without a visual designing tools. The basics of SILK are the tasks and relations (rules) between them. It is similar with the SCUFL language, but not relying on XML in order to allow the introduction of more workflow specific issues. Moreover, an event-condition-action (ECA) approach allows a greater flexibility when expressing data and task dependencies, as well as the creation of adaptive workflows which can react to changes in the configuration of the Grid or in the workflow itself. Changes inside the grid are handled by creating specific rules which allow resource selection based on various task scheduling criteria. Modifications of the workflow are usually accomplished either by inserting or retracting at runtime rules belonging to it or by modifying the executor of the task in case a better one is found. The former implies changes in its structure while the latter does not necessarily mean changes of the resource but more precisely changes of the algorithm used for solving the task. More details can be found in [3].

Another important platform component is the data indexing and storage service, GDIS, providing features for data storage, indexing data using a specialized RDBMS, finding data by various conditions, querying external services and keeping track of temporary data generated by other components. The data storage component part of GDIS is responsible for storing the data by using available storage backends such as local disk file systems (ext3), local cluster storage (GFS) or distributed file systems (HDFS). A front-end GridFTP service is capable of interacting with the storage domains on behalf of the clients and in a uniform way and also enforces the security restrictions provided by other specialized services and related with data access. The data indexing is performed by PostGIS. An advanced and flexible interface for searching the project's geographical repository is built around a

custom query language (LLQL - Lisp Like Query Language) designed to provide fine grained access to the data in the repository and to query external services (e.g. for exploiting the connection with GENESI-DR catalog). More details can be found in [4].

The Workload Management System (WMS) provides two types of resource managers. The first one will be based on Condor HTC and use Condor as a job manager for task dispatching and working nodes (for development purposes) while the second one will use GT4 GRAM (for production purposes). The WMS main component, the Grid Task Dispatcher (GTD), is responsible for the interaction with other internal services as the composition engine in order to facilitate access to the processing platform. Its main responsibilities are to receive tasks from the workflow engine or directly from user interface, to use a task description language (the ClassAd meta language in case of Condor HTC) for job units, to submit and check the status of jobs inside the workload management system and to retrieve job logs for debugging purposes. More details can be found in [4].

A particular component of the platform is eGLE, the eLearning environment. It provides the functionalities necessary to create the visual appearance of the lessons through the usage of visual containers like tools, patterns and templates. The teacher uses the platform for testing the already created lessons, as well as for developing new lesson resources, such as new images and workflows describing graph-based processing. The students execute the lessons or describe and experiment with new workflows or different data. The eGLE database includes several workflow-based lesson descriptions, teaching materials and lesson resources, selected satellite and spatial data. More details can be found in [5].

A first training event of using the platform was organized in September 2009 during 11th SYNASC symposium (links to the demos, testing interface, and exercises are available on project site [1]). The eGLE component was presented at 4th GPC conference in May 2009. Moreover, the functionality of the platform will be presented as demo in April 2010 at 5th EGEE User forum.

References:

[1] GiSHEO consortium, Project site, http://gisheo.info.uvt.ro

[2] D. Petcu, D. Zaharie, M. Neagul, S. Panica, M. Frincu, D. Gorgan, T. Stefanut, V. Bacu, Remote Sensed Image Processing on Grids for Training in Earth Observation. In Image Processing, V. Kordic (ed.), In-Tech, January 2010.

[3] M. Neagul, S. Panica, D. Petcu, D. Zaharie, D. Gorgan, Web and Grid Services for Training in Earth Observation, IDAACS 2009, IEEE Computer Press, 241-246

[4] M. Frincu, S. Panica, M. Neagul, D. Petcu, Gisheo: On Demand Grid Service Based Platform for EO Data Processing. HiperGrid 2009, Politehnica Press, 415–422.

[5] D. Gorgan, T. Stefanut, V. Bacu, Grid Based Training Environment for Earth Observation, GPC 2009, LNCS 5529, 98-109