PROCESSING REMOTE SENSING IMAGES ON A GRID-BASED PLATFORM

Silviu Panica, Marian Neagul and Dana Petcu
Computer Science Department, West University of Timisoara
B-dul Vasile Parvan 4, 300223 Timisoara, Romania

ABSTRACT

Huge quantity of remote sensing data is acquired daily by several satellites and only few of them are really exploited. The public availability of a considerable part of these data allows the development of new innovative applications. A minimal training in processing remote sensing images is needed to develop such applications. Unfortunately, training the developers of Earth observation applications is currently addressed only by few institutions and platforms. In this context, we have built recently a training platform distinguished from other solutions by its high degree of interactivity and extensibility. The fast response to the multiple users' processing requests is possible due to the usage of Grid computing technologies that is hidden behind a friendly Web-accessible interface shortly described in this paper.

KEYWORDS

Grid computing, Image processing, Earth observation, Distributed data management.

1. INTRODUCTION

Due to its intensive data processing and highly distributed organization, the multidisciplinary Earth Observation (EO) applications community is highly interested in the exploitation of Grid computing infrastructures. Several petabytes of already acquired data are presently underexploited due to the fact that getting the results in a reasonable time requires more computing power that currently is available in the biggest data centers. If these data are made publicly available, even with the support of Grid technologies mainly focusing on computing facilities, an efficient distributed infrastructure to handle and treat very large data sets is still missing. In order to facilitate the access to data, their processing and visualization, a special attention was given in the last decade to the development of Web portals integrating standard and EO specific services. In few cases these portals expose Web or Grid-based computational services or Web access to parallel processing facilities. Despite the availability of these new Internet-based remote facilities, human resources involved in the development of new services or exploitation of the existing ones are suffering from low level of training either in Earth observation techniques, either in using the new technologies.

In this context we proposed recently a Grid-based training platform for Earth Observation, namely GiSHEO. Its design and different views of particular platform components functionality were already reported last year (Neagul et al, Petcu et al, and Petcu 2009). Following the results of stressing the platform under real usage conditions in classroom trainings in the academic year 2009-2010, the requirements for the user interface were modified and the development of a new Web user interface was necessary affecting partially also the services interfaces. In this paper we report shortly these changes.

2. SHORT OVERVIEW OF THE GISHEO PLATFORM

GiSHEO's architecture is basically a service oriented one. The Grid-enabled platform for satellite image processing is structured on several levels including user, security, service, processing and a data level. The user level is in charge with the access to the web user interface (built by using Google's Web Toolkit framework). The security level provides security context for both users and services. The service level

exposes internal mechanisms belonging to the GiSHEO platform by using various Web services technologies such as the followings: EO services – processing applications are represented through a Web service interface; the workflow service – internal workflow engine which can be accessed by using a specialized Web service; data indexing and discovery services – access to the GiSHEO's data management mechanisms. At processing level the GiSHEO platform proposes two models for data processing by either using Condor HTC, a direct job submission using Condor's specific Web services, or Globus Toolkit 4 through GRAM. At data level we have the datasets database which contains the satellite imagery repository and processing application datasets used by applications to manipulate satellite images. At the date of this paper the repository includes authorized copies of NASA public available remote sensing images, photograms specific for the geographical region of the developers, as well as connections with ESA's GENESI-DR catalogue through a particular Web service.

The GiSHEO processing platform consists of two parts, the interface exposed as a Web service (WS) and the workload management system. The interface is built by using AXIS2 Web Service technology and is responsible for the interaction with other internal services as the Gisheo Data Index Service (GDIS) in order to facilitate access to the processing platform. Its main responsibilities are at this point to receive tasks from the workflow engine or directly from user interface, to use a task description language (the ClassAd meta language for example in case of Condor HTC) in order to describe a job unit, to submit and check the status of jobs inside the workload management system and to retrieve job logs for debugging purposes. As a workload management system, GiSHEO uses Condor HTC and provides Condor HTC resource manager accessible through it built-in Web service interface to access an internal component called Condor Job Manager used for task dispatching and administration.

The Web-based user interface is designed as a client to the platform services. New user interfaces or applications can be easily build on top of the existing Web and Grid services that are publicly exposed on the project Web site. Due to the fact that EO applications are data-intensive, the key element in any Web portal for EO is the selection of the data and, only after it, the selection of the processing that will be applied to them. Note GiSHEO's current solution in Figure 1: the data information is central; each data has associated a list of tasks that can be launched using it depending on its type.

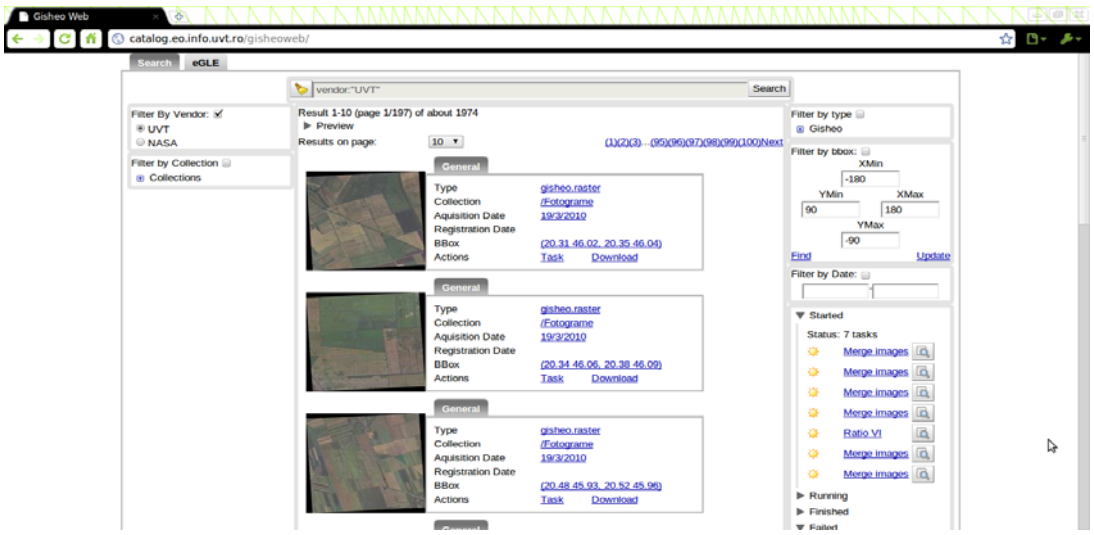

Figure 1. GiSHEO's Web-based interface - front page: photogram catalog

The EO data selection in different EO portals range from simple selection from list based catalogues to visual selection of region of interests. The early interface of GiSHEO has been designed having in mind a low degree of knowledge about data types and therefore the visual selection was preferred (see the interface functionality in the demos provided on the project website). But this option has put a high stress on the platform, as well as on the user's Web browser, in order to be able to present multiple pre-views of the huge data (at different zooming scales) available for the same region simultaneous. Moreover the computing task dependability on the type of the input data was hardly solved in the case of multiple inputs representing data

from the same region of interest. The new solution that is proposed allows the user to select the images to process using the location, type, date and so on parameters, and presents the data available in a list form, each entry having only one preview form and a list of possible tasks to be applied. The user can specify a particular location in which he or she is interested. More complex type filtering is also available. A specific language was designed for these kinds of selections and can be used when someone wants to build his/her user interface to the platform – documentation is available on project site.

Note that the input and output (huge) images are not transferred to the user site, only at request, and all the processing are taking place where the data are, on GiSHEO platform. The processing results are stored at remote site in catalogues that are specified by the user and which can be shared later one with other users. Since the processing time of the huge satellite images can be of seconds order (depending on the task type), the communication with the user is asynchronous. After the task completion, the user can retrieve the results.

As mentioned earlier, the platform is designed with the training purpose in mind. The easy development of the EO lessons based on the interactive experimental facilities of the platform has been also a clear goal of the project. The eGLE component of the platform for e-learning allows the teacher to build, in a short time, new lessons based on templates for texts, images, videos etc.

3. RELATED WORK

GiSHEO follows after considerable world-wide efforts in Grid-based experiments for Earth observation. We consider that two periods can be distinguished in Grid-based Earth observation developments. The first one is that of early experiments and proof-of-concepts finalized by European DataGrid, SARA Digital Puglia or GEOGrid projects. The first period finished with the studies delivered by the European projects D4Science and DEGREE about the challenges that the Earth Sciences are imposing on Grid infrastructure, as well as several case studies in which Grid are useful of production environments. The second period is the one of the production environments. The platform called Grid Processing On Demand, shortly G-POD currently offers a Grid-based platform for remote processing of the satellite images provided by ESA and offers several satellite image processing services for environmental management. Moreover, GENESI-DR offers an interface for digital data discovery and retrieval, while raw data are processed using G-POD facilities. The Landsat Grid Prototype LGP is using Grid computing to generate single scenes from the composite of multiple input scenes. EGEE-3 and SEE-Grid-SCI e-infrastructures projects have build environmental applications based on satellite data including also some of the ones provided by GENESI-DR. Note that the GiSHEO's eGLE component is connected to the SEE-Grid-SCI platform, and the catalog to GENESI-DR.

The closest solution to GiSHEO's one is G-POD. Both have Web services interfaces to access the platform facilities. Opposite to G-POD that is designed for commercial aims and offers very complex processing services on proprietary images, GiSHEO is designed for training aims and the processing tasks are simple and operate on public available data. Complex applications can be build on top of GiSHEO platform, to reach the level of complexity of G-POD services, only by using its workflow based services that are exposed and by exploiting a pre-knowledge of the user in Earth observation processing requirements.

Currently there is only a few number of resources involved in educational activities in EO. One of the most complex is EduSpace. While the training material is more consistent than that of the GiSHEO platform, the interactivity and flexibility degree in the relationship platform-user is lower compared with GiSHEO solution.

REFERENCES

Neagul, M., Panica, S., Petcu, D., Zaharie, D. and Gorgan D., 2009. Web and Grid Services for Training in Earth Observation, *Proceedings of IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*. Rende, Italy, pp. 241-246.

Petcu, D., 2009. Challenges of Data Processing for Earth Observation in Distributed Environments, *Intelligent Distributed Computing III, Studies in Computational Intelligence* SCI 237, Springer, pp. 9-19.

Petcu, D., Zaharie, D., Neagul, M., Panica, S., Frincu, M., Gorgan, D., Stefanut, T. and Bacu, V., 2009. Remote Sensed Image Processing on Grids for Training in Earth Observation. *Image Processing,* In-Tech, Vienna, pp. 115 – 140.