



gProcess and ESIP Platforms for Satellite Imagery Processing over the Grid

Victor Bacu (1), Dorian Gorgan (1), Denisa Rodila (1), Florin Pop (2), Gabriel Neagu (3), and Dana Petcu (4)

(1) Technical University of Cluj-Napoca, Computer Science Department, Cluj-Napoca, Romania (dorian.gorgan@cs.utcluj.ro), (2) Politehnica University of Bucharest, Bucharest, Romania (florinpop@cs.pub.ro), (3) National Research Institute of Informatics, Bucharest, Romania (gneagu@ici.ro), (4) Western University of Timisoara, Timisoara, Romania (petcu@info.uvt.ro)

The Environment oriented Satellite Data Processing Platform (ESIP) is developed through the SEE-GRID-SCI (SEE-GRID eInfrastructure for regional eScience) co-funded by the European Commission through FP7 [1]. The gProcess Platform [2] is a set of tools and services supporting the development and the execution over the Grid of the workflow based processing, and particularly the satellite imagery processing. The ESIP [3], [4] is build on top of the gProcess platform by adding a set of satellite image processing software modules and meteorological algorithms.

The satellite images can reveal and supply important information on earth surface parameters, climate data, pollution level, weather conditions that can be used in different research areas. Generally, the processing algorithms of the satellite images can be decomposed in a set of modules that forms a graph representation of the processing workflow. Two types of workflows can be defined in the gProcess platform: abstract workflow (PDG – Process Description Graph), in which the user defines conceptually the algorithm, and instantiated workflow (iPDG – instantiated PDG), which is the mapping of the PDG pattern on particular satellite image and meteorological data [5]. The gProcess platform allows the definition of complex workflows by combining data resources, operators, services and sub-graphs.

The gProcess platform is developed for the gLite middleware that is available in EGEE and SEE-GRID infrastructures [6]. gProcess exposes the specific functionality through web services [7]. The Editor Web Service retrieves information on available resources that are used to develop complex workflows (available operators, sub-graphs, services, supported resources, etc.). The Manager Web Service deals with resources management (uploading new resources such as workflows, operators, services, data, etc.) and in addition retrieves information on workflows. The Executor Web Service manages the execution of the instantiated workflows on the Grid infrastructure. In addition, this web service monitors the execution and generates statistical data that are important to evaluate performances and to optimize execution. The Viewer Web Service allows access to input and output data.

To prove and to validate the utility of the gProcess and ESIP platforms there were developed the GreenView and GreenLand applications. The GreenView related functionality includes the refinement of some meteorological data such as temperature, and the calibration of the satellite images based on field measurements. The GreenLand application performs the classification of the satellite images by using a set of vegetation indices. The gProcess and ESIP platforms are used as well in GiSHEO project [8] to support the processing of Earth Observation data over the Grid in eGLE (GiSHEO eLearning Environment).

Experiments of performance assessment were conducted and they have revealed that the workflow-based execution could improve the execution time of a satellite image processing algorithm [9]. It is not a reliable solution to execute all the workflow nodes on different machines. The execution of some nodes can be more time consuming and they will be performed in a longer time than other nodes. The total execution time will be affected because some nodes will slow down the execution. It is important to correctly balance the workflow nodes. Based

on some optimization strategy the workflow nodes can be grouped horizontally, vertically or in a hybrid approach. In this way, those operators will be executed on one machine and also the data transfer between workflow nodes will be lower.

The dynamic nature of the Grid infrastructure makes it more exposed to the occurrence of failures. These failures can occur at worker node, services availability, storage element, etc. Currently gProcess has support for some basic error prevention and error management solutions. In future, some more advanced error prevention and management solutions will be integrated in the gProcess platform.

References

- [1] SEE-GRID-SCI Project, <http://www.see-grid-sci.eu/>
- [2] Bacu V., Stefanut T., Rodila D., Gorgan D., Process Description Graph Composition by gProcess Platform. HiPerGRID - 3rd International Workshop on High Performance Grid Middleware, 28 May, Bucharest. Proceedings of CSCS-17 Conference, Vol.2., ISSN 2066-4451, pp. 423-430, (2009).
- [3] ESIP Platform, http://wiki.egee-see.org/index.php/JRA1_Commonalities
- [4] Gorgan D., Bacu V., Rodila D., Pop Fl., Petcu D., Experiments on ESIP - Environment oriented Satellite Data Processing Platform. SEE-GRID-SCI User Forum, 9-10 Dec 2009, Bogazici University, Istanbul, Turkey, ISBN: 978-975-403-510-0, pp. 157-166 (2009).
- [5] Radu, A., Bacu, V., Gorgan, D., Diagrammatic Description of Satellite Image Processing Workflow. Workshop on Grid Computing Applications Development (GridCAD) at the SYNASC Symposium, 28 September 2007, Timisoara, IEEE Computer Press, ISBN 0-7695-3078-8, 2007, pp. 341-348 (2007).
- [6] Gorgan D., Bacu V., Stefanut T., Rodila D., Mihon D., Grid based Satellite Image Processing Platform for Earth Observation Applications Development. IDAACS'2009 - IEEE Fifth International Workshop on "Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications", 21-23 September, Cosenza, Italy, IEEE Published in Computer Press, 247-252 (2009).
- [7] Rodila D., Bacu V., Gorgan D., Integration of Satellite Image Operators as Workflows in the gProcess Application. Proceedings of ICCP2009 - IEEE 5th International Conference on Intelligent Computer Communication and Processing, 27-29 Aug, 2009 Cluj-Napoca. ISBN: 978-1-4244-5007-7, pp. 355-358 (2009).
- [8] GiSHEO consortium, Project site, <http://gisheo.info.uvt.ro>
- [9] Bacu V., Gorgan D., Graph Based Evaluation of Satellite Imagery Processing over Grid. ISPDC 2008 - 7th International Symposium on Parallel and Distributed Computing, July 1-5, 2008, Krakow, Poland. IEEE Computer Society 2008, ISBN: 978-0-7695-3472-5, pp. 147-154.