



Flexible Description Language for HPC based Processing of Remote Sense Data

Constantin Nandra, Dorian Gorgan, and Victor Bacu

Technical University of Cluj-Napoca, Computer Science Department, Cluj-Napoca, Romania (dorian.gorgan@cs.utcluj.ro)

When talking about Big Data, the most challenging aspect lays in processing them in order to gain new insight, find new patterns and gain knowledge from them. This problem is likely most apparent in the case of Earth Observation (EO) data. With ever higher numbers of data sources and increasing data acquisition rates, dealing with EO data is indeed a challenge [1]. Geoscientists should address this challenge by using flexible and efficient tools and platforms. To answer this trend, the BigEarth project [2] aims to combine the advantages of high performance computing solutions with flexible processing description methodologies in order to reduce both task execution times and task definition time and effort. As a component of the BigEarth platform, WorDeL (Workflow Description Language) [3] is intended to offer a flexible, compact and modular approach to the task definition process.

WorDeL, unlike other description alternatives such as Python or shell scripts, is oriented towards the description topologies, using them as abstractions for the processing programs. This feature is intended to make it an attractive alternative for users lacking in programming experience.

By promoting modular designs, WorDeL not only makes the processing descriptions more user-readable and intuitive, but also helps organizing the processing tasks into independent sub-tasks, which can be executed in parallel on multi-processor platforms in order to improve execution times.

As a BigEarth platform [4] component, WorDeL represents the means by which the user interacts with the system, describing processing algorithms in terms of existing operators and workflows [5], which are ultimately translated into sets of executable commands. The WorDeL language has been designed to help in the definition of compute-intensive, batch tasks which can be distributed and executed on high-performance, cloud or grid-based architectures in order to improve the processing time.

Main references for further information:

[1] Gorgan, D., "Flexible and Adaptive Processing of Earth Observation Data over High Performance Computation Architectures", International Conference and Exhibition Satellite 2015, August 17-19, Houston, Texas, USA.

[2] Bigearth project - flexible processing of big earth data over high performance computing architectures. <http://cgis.utcluj.ro/bigearth>, (2014)

[3] Nandra, C., Gorgan, D., "Workflow Description Language for Defining Big Earth Data Processing Tasks", Proceedings of the Intelligent Computer Communication and Processing (ICCP), IEEE-Press, pp. 461-468, (2015).

[4] Bacu, V., Stefan, T., Gorgan, D., "Adaptive Processing of Earth Observation Data on Cloud Infrastructures Based on Workflow Description", Proceedings of the Intelligent Computer Communication and Processing (ICCP), IEEE-Press, pp.444-454, (2015).

[5] Mihon, D., Bacu, V., Colceriu, V., Gorgan, D., "Modeling of Earth Observation Use Cases through the KEOPS System", Proceedings of the Intelligent Computer Communication and Processing (ICCP), IEEE-Press, pp. 455-460, (2015).