
A2S – Processing services for the rapid exploitation of satellite data streams on HPC platforms (mésocentre de Strasbourg)

David Michéa*¹

¹Université Strasbourg – université de Strasbourg – France

Résumé

Since more than 20 years, "Earth Observation" (EO) satellites have provided a wealth of data, and currently routine monitoring of our environment at the global scale, is becoming possible through the delivery of this unprecedented amount of data. This growing volume of data poses a major challenge to achieve its full potential in terms of data exploitation, especially for a rapid processing.

In this context, the University of Strasbourg supports the A2S 'Alsace Aval Sentinel' programme which aims at building an operative processing environment to maximize the exploitation of EO data. This initiative is also part of the national data repository initiatives (THEIA Land, ForM@Ter) designed to foster the use of satellite images for the monitoring of land surfaces and the solid earth, through regional and thematic scientific competence centres.

The presentation focuses on the presentation of the distributed processing chains which are currently being implemented on the high performance computing facility of University of Strasbourg (Mésocentre Unistra). The innovative processing methods being implemented consist, among other, of:

- a chain for the quantification of earth surface deformation from several processes (tectonics, landslides, glaciers) and making use of optical and SAR images;
- a chain for the quantification of urban (and more generally impermeable surfaces) developments;

The system is designed for both near-real time routine processing of images within one day after acquisition, and on-demand processing of image archives (past and current satellite missions). The processing chains are currently designed for processing large areas of 1000 x 1000 km, and will be extended to 5000 x 5000 km in a second phase.

The infrastructure (hardware/software) is based on a HPC platform. Indeed, several implementations need a large amount of memory and a high level of parallelism at the node level since they are not implemented for running on distributed memory architectures. The total amount of computation and the diversity of the algorithms, even in a context of embarrassingly parallel computations, is sufficient to justify the use of such a platform.

*Intervenant

The workflow management system (WFMS) Fireworks is used to arrange, monitor and evaluate the performance of the different sequences of tasks. Fireworks has been chosen because it interfaces natively with HPC's queuing systems. As several operations require a lot of data and are limited by the disk I/O bandwidth, a high bandwidth between the HPC cluster and the storage system is implemented. The scratch disks, installed on the computing nodes, are used to build a high-bandwidth distributed file-system by connecting them with a fast low-latency network. This file-system acts as a cache for the high capacity storage system.