Challenges in building high performance geoscientific spatial data infrastructures

Fabrice Dubros, Agnes Tellez-Arenas, Faiza Boulahya, Robin Quique, Goneri Le Cozanne, and Hideo Aochi
BRGM, 3 avenue Cloude Guillemin, 45060 Orléans

One of the main challenges in Geosciences is to deal with both the huge amounts of data available nowadays and the increasing need for fast and accurate analysis.

On one hand, computer aided decision support systems remain a major tool for quick assessment of natural hazards and disasters. High performance computing lies at the heart of such systems by providing the required processing capabilities for large three-dimensional time-dependent datasets.

On the other hand, information from Earth observation systems at different scales is routinely collected to improve the reliability of numerical models. Therefore, various efforts have been devoted to design scalable architectures dedicated to the management of these data sets (Copernicus, EarthCube, EPOS). Indeed, standard data architectures suffer from a lack of control over data movement. This situation prevents the efficient exploitation of parallel computing architectures as the cost for data movement has become dominant.

In this work, we introduce a scalable architecture that relies on high performance components. We discuss several issues such as three-dimensional data management, complex scientific workflows and the integration of high performance computing infrastructures. We illustrate the use of such architectures, mainly using off-the-shelf components, in the framework of both coastal flooding assessments and earthquake early warning systems.