Geophysical Research Abstracts
Vol. 14, EGU2012-9543, 2012
EGU General Assembly 2012

The Virtual Earthquake and Seismology Research Community e-science environment in Europe (VERCE) FP7-INFRA-2011-2 project

J.-P. Vilotte (1), M. Atkinson (2), A. Michelini (3), H. Igel (4), and T. van Eck (5)

(1) Institut de Physique du Globe de Paris, France (vilotte@ipgp.fr), (2) School of Informatics, University of Edinburgh, UK, (3) INGV, Roma, Italy, (4) Ludwig-Maximilians-Universitaet, Germany, (5) ORFEUS, KNMI, Netherland

Increasingly dense seismic and geodetic networks are continuously transmitting a growing wealth of data from around the world. The multi-use of these data leaded the seismological community to pioneer globally distributed open-access data infrastructures, standard services and formats, e.g., the Federation of Digital Seismic Networks (FDSN) and the European Integrated Data Archives (EIDA).

Our ability to acquire observational data outpaces our ability to manage, analyze and model them. Research in seismology is today facing a fundamental paradigm shift. Enabling advanced data-intensive analysis and modeling applications challenges conventional storage, computation and communication models and requires a new holistic approach. It is instrumental to exploit the cornucopia of data, and to guarantee optimal operation and design of the high-cost monitoring facilities.

The strategy of VERCE is driven by the needs of the seismological data-intensive applications in data analysis and modeling. It aims to provide a comprehensive architecture and framework adapted to the scale and the diversity of those applications, and integrating the data infrastructures with Grid, Cloud and HPC infrastructures. It will allow prototyping solutions for new use cases as they emerge within the European Plate Observatory Systems (EPOS), the ESFRI initiative of the solid Earth community.
Computational seismology, and information management, is increasingly revolving around massive amounts of data that stem from: (1) the flood of data from the observational systems; (2) the flood of data from large-scale simulations and inversions; (3) the ability to economically store petabytes of data online; (4) the evolving Internet and Data-aware computing capabilities. As data-intensive applications are rapidly increasing in scale and complexity, they require additional services-oriented architectures offering a virtualization-based flexibility for complex and re-usable workflows.

Scientific information management poses computer science challenges: acquisition, organization, query and visualization tasks scale almost linearly with the data volumes. Commonly used FTP-GREP metaphor allows today to scan gigabyte-sized datasets but will not work for scanning terabyte-sized continuous waveform datasets. New data analysis and modeling methods, exploiting the signal coherence within dense network arrays, are nonlinear. Pair-algorithms on N points scale as N2. Wave form inversion and stochastic simulations raise computing and data handling challenges These applications are unfeasible for tera-scale datasets without new parallel algorithms that use near-linear processing, storage and bandwidth, and that can exploit new computing paradigms enabled by the intersection of several technologies (HPC, parallel scalable database crawler, data-aware HPC).

This issues will be discussed based on a number of core pilot data-intensive applications and use cases retained in VERCE. This core applications are related to: (1) data processing and data analysis methods based on correlation techniques; (2) cpu-intensive applications such as large-scale simulation of synthetic waveforms in complex earth systems, and full waveform inversion and tomography. We shall analyze their workflow and data flow, and their requirements for a new service-oriented architecture and a data-aware platform with services and tools. Finally, we will outline the importance of a new collaborative environment between seismology and computer science, together with the need for the emergence and the recognition of 'research technologists' mastering the evolving data-aware technologies and the data-intensive research goals in seismology.