The VERCE platform: Enabling Computational Seismology via Streaming Workflows and Science Gateways. A Forecasting Model Testing Centre

Geophysical Research Abstracts
Vol. 17, EGU2015-6129-8, 2015
EGU General Assembly 2015

The VERCE platform: Enabling Computational Seismology via Streaming Workflows and Science Gateways

Alessandro Spinuso (1), Rosa Filgueira (2), Amrey Krause (2), Jonas Matser (1), Emanuele Casarotti (3), Federica Magnoni (3), Andre Gemund (4), Laurent Frobert (5), Lion Krischer (6), and Malcolm Atkinson (2)

(1) ORFEUS-KNMI, R&D Observation an Data Technology, Utrecht, Netherlands, (2) University of Edinburgh, UK, (3) INGV - Italian National Institute of Geophysics and Vulcanology, Italy, (4) SCAI - Fraunhofer, Germany, (5) EMSC-CSEM, Euro-Med Seismological Centre, France, (6) LMU - Ludwig Maximilians-Universiteit, Germany

The VERCE project is creating an e-Science platform to facilitate innovative data analysis and coding methods that fully exploit the wealth of data in global seismology. One of the technologies developed within the project is the Dispel4Py python library, which allows to describe abstract stream-based workflows for data-intensive applications and to execute them in a distributed environment. At runtime Dispel4Py is able to map workflow descriptions dynamically onto a number of computational resources (Apache Storm clusters, MPI powered clusters, and shared-memory multi-core machines, single-core machines), setting it apart from other workflow frameworks. Therefore, Dispel4Py enables scientists to focus on their computation instead of being distracted by details of the computing infrastructure they use.

Among the workflows developed with Dispel4Py in VERCE, we mention here those for Seismic Ambient Noise Cross-Correlation and MISFIT calculation, which address two data-intensive problems that are common in computational seismology. The former, also called Passive Imaging, allows the detection of relative seismic-wave velocity variations during the time of recording, to be associated with the stress-field changes that occurred in the test area. The MISFIT instead, takes as input the synthetic seismograms generated from HPC simulations for a certain Earth model and earthquake and, after a preprocessing stage, compares them with real observations in order to foster subsequent model updates and improvement (Inversion).

The VERCE Science Gateway exposes the MISFIT calculation workflow as a service, in combination with the simulation phase. Both phases can be configured, controlled and monitored by the user via a rich user interface which is integrated within the gUSE Science Gateway framework, hiding the complexity of accessing third parties data services, security mechanisms and enactment on the target resources. Thanks to a modular extension to the Dispel4Py framework, the system collects provenance data adopting the W3C-PROV data model. Provenance recordings can be explored and analysed at run time for rapid diagnostic and workflow steering, or later for further validation and comparisons across runs. We will illustrate the interactive services of the gateway and the capabilities of the produced metadata, coupled with the VERCE data management layer based on iRODS.

The Cross-Correlation workflow was evaluated on SuperMUC, a supercomputing cluster at the Leibniz Supercomputing Centre in Munich, with 155,656 processor cores in 9400 compute nodes. SuperMUC is based on the Intel Xeon architecture consisting of 18 Thin Node Islands and one Fat Node Island. This work has only had access to the Thin Node Islands, which contain Sandy Bridge nodes, each having 16 cores and 32 GB of memory. In the evaluations we used 1000 stations, and we applied two types of methods (whiten and non-whiten) for pre-processing the data. The workflow was tested on a varying number of cores (16, 32, 64, 128, and 256 cores) using the MPI mapping of Dispel4Py. The results show that Dispel4Py is able to improve the performance by increasing the number of cores without changing the description of the workflow.