GC11-solidearth-7, updated on 29 Nov 2023
https://doi.org/10.5194/egusphere-gc11-solidearth-7
Galileo Conference: Solid Earth and Geohazards in the Exascale Era

BackTrackBB workflow for seismic source detection and location with PyCOMPSs parallel computational framework

Natalia Poiata[1,2], Javier Conejero[3], Rosa M. Badia[3], and Jean-Pierre Vilotte[4]
[1]National Institute for Earth Physics, Măgurele, Romania (natalia@infp.ro)
[2]International Seismological Centre, Thatcham, UK
[3]Department of Computer Sciences, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain
[4]Université Paris Cité, Institut de Physique du Globe de Paris, CNRS-INSU, Paris, France

In this work we present a scalable parallelization with PyCOMPSs (Tejedor et al., 2017; the Python binding of COMPSs) of the Python-based workflow BackTrackBB (Poiata et al., 2016) for the automatic detection and location of seismic sources using continuous waveform data recorded by regular to large seismic networks. PyCOMPSs is a task-based programming model for Python applications that relies in a powerful runtime able to extract dynamically the parallelism among tasks and execute them in distributed environments (e.g., HPC Clusters, Cloud infrastructures, etc.) transparently to the users. BackTrackBB with PyCOMPSs implementation allows to fully parallelize the seismic source detection and location process making it efficient and portable in terms of the use of available HPC resources.

We provide details of the BackTrackBB workflow implementation with PyCOMPSs and discuss its performance by presenting the results of the scalability tests and memory usage analysis. All the tests have been performed on the MareNostrum4 High-Performance computer of the Barcelona Supercomputing Centre. The first version of the BackTrackBB with PyCOMPSs workflow was developed in the context of the European Centre Of Excellence (CoE) ChEESE for Exascale computing in solid earth sciences. The initial workflow developments and performance tests made use of a simplified synthetic dataset emulating a large-scale seismic network deployment in a seismically active area and corresponding to 100 vertical sensors recording a month of continuous waveforms at a sampling rate of 100 sps. In the following testing step, the workflow was applied to the real-case two-month long dataset from Vrancea seismic region in Romania (corresponding to the 60-190 km deep earthquakes activity). Real seismic data scenario proved to present some challenges in terms of the data-quality control, that often occurs in the case of continuous waveforms recorded by the seismic observatories. This issue have been resolved and corresponding modifications were included in the final version of BackTrackBB with PyCOMPSs. The real dataset tests showed that the workflow allows improved detection and location of seismic events through the efficient processing of the large continuous seismic data with important performance and scalability improvements.

We show that BackTrackBB with PyCOMPSs workflow enables generation of fully reproducible, seismic catalogues (or seismic catalogues realizations) through the analysis of the continuous large (in terms of the number of seismic stations, data record length and covered area) seismic data-sets. Such implementations making use of advances full-waveform detection and location

methods are currently highly-challenging or, some-times, impossible due to the amount of required main memory or unfeasible time to solution. PyCOMPSs has demonstrated to be able to deal with both issues successfully allowing to explore in greater depth the usage with BackTrackBB method. Workflows such as BackTrackBB with PyCOMPSs has the ability to significantly improve the detections and location process that is currently in place at seismological observatories or network operation centres, providing fully reproducing detailed catalogues in the seismically-active regions and allowing multiple input parameters testing (e.g., station configuration, velocity models).