

Distributed trace processing on Cloud resources

Monday, 19 May 2014 16:45 (15 minutes)

The EU-funded project VERCE (<http://verce.eu/>) aims to address specific seismological use-cases employing resources spanning available e-Infrastructures on the basis of requirements elicited from the seismology community. It provides a service-oriented infrastructure to deal with the challenges the researchers face in the data-intensive and high-performance computations employed in modern seismology. In particular, the implementation is driven by two major use-cases. The first is the computationally intensive forward and inverse modelling of Earth system models, which is implemented with support for multiple waveform simulators running on HPC systems and x86 clusters, and the data-oriented seismic wave cross-correlation.

In this talk we present a service that is intended to complement the services of the VERCE projects infrastructure, a service that utilizes Cloud resources to offer simplified execution of data pre- and post-processing scripts. It offers users access to the ObsPy seismological toolbox for processing data with the Python programming language, executed on virtual Cloud resources in a secured sandbox.

The solution encompasses a frontend with a modern graphical user interface, a messaging infrastructure as well as Python worker nodes for background processing. All components are deployable in the Cloud and have been tested on different environments based on OpenStack and OpenNebula. Deployments on commercial, public Clouds will be tested in the future.

Wider impact and conclusions

Once the development and subsequent testing have been completed, its planned to offer the service running the EGI Fedcloud and usable by members of the VERCE.eu EGI VOs.

URL(s) for further info

<http://www.verce.eu/>

Description of work

The EU-funded project VERCE (<http://verce.eu/>) aims to address specific seismological use-cases employing resources spanning available e-Infrastructures on the basis of requirements elicited from the seismology community. It provides a service-oriented infrastructure to deal with the challenges the researchers face in the data-intensive and high-performance computations employed in modern seismology. In particular, the implementation is driven by two major use-cases. The first is the computationally intensive forward and inverse modelling of Earth system models, which is implemented with support for multiple waveform simulators running on HPC systems and x86 clusters, and the data-oriented seismic wave cross-correlation.

In this talk we present a service that is intended to complement the services of the VERCE projects infrastructure, a service that utilizes Cloud resources to offer simplified execution of data pre- and post-processing scripts. It offers users access to the ObsPy seismological toolbox for processing data with the Python programming language, executed on virtual Cloud resources in a secured sandbox.

The solution encompasses a frontend with a modern graphical user interface, a messaging infrastructure as well as Python worker nodes for background processing. All components are deployable in the Cloud and have been tested on different environments based on OpenStack and OpenNebula. Deployments on commercial, public Clouds will be tested in the future.

The frontend GUI provides graphical query tools to find suitable input traces. For this purpose, data centres with FDSN compliant Web services can be queried. Based on the users data and parameter selection, the frontend populates preconfigured/user-editable python scripts. With the help of a high-throughput/light-weight messaging service (Beanstalkd), these python jobs are distributed to available worker nodes/VMs.

Primary authors: GEMUEND, Andre (FRAUNHOFER); KRISCHER, Lion (Ludwig-Maximilians-Universität

München); Mr CLAUS, Steffen (Fraunhofer SCAI)

Presenters: GEMUEND, Andre (FRAUNHOFER); Mr CLAUS, Steffen (Fraunhofer SCAI)

Session Classification: e-Infrastructure Services for Earth Science

Track Classification: Success stories in using e-Infrastructures for research (Track Leaders: E. Karagkou, P. Castejon)