D-NA2.3: Application demonstrators for dissemination and public outreach

30/10/2013

Project acronym:	VERCE
Project n°:	283543
Funding Scheme:	Combination of CP & CSA
Call Identifier:	FP7-INFRASTRUCTURES-2011-2
WP:	WP2/NA2, Pilot applications and use cases
Filename:	D-NA2.3.1.pdf
Author(s):	A. Michelini
Location:	http://www.verce.eu/Repository/Deliverables/RP2/
Type of document:	Deliverable
Dissemination level:	Public
Status:	Draft
Due date of delivery:	30/10/ 2013
Reviewer:	J.-P. Vilotte
Keywords:	data-intensive, cpu-intensive, HPC, earthquake, seismology, data infrastructure, forward modeling, inversion

Version	*Author*	*Date*	*Comments*
1	A. Michelini (INGV)	27/10/2013	Initial draft for comments
2	J-P Vilotte (IPGP)	30/10/2012	Revision

Copyright notice

Contents

Executive Summary

The main objectives of WP2/NA2 are: (1) select existing pilot data-intensive applications and design sound use case scenarios; (2) analyze and define a use case implementation strategy during the project with WP8, WP7 and WP9; (3) support and evaluate the "productising" transition of the methods and their implementation performed by WP8; (4) support and evaluate the deployment and the efficiency of the pilot applications and their use case scenarios on the VERCE platform; (5) define in collaboration with NA3 documentation and tailored training session material; (6) provide requirements and support to WP7 and WP9 for tailored interfaces of the scientific gateways targeted to the developers and the users.

This report of the WP2/NA2 should concern the demonstrators for dissemination and public outreach. The demonstrators so far developed, however, are primarily proof-of-concept of the VERCE platform. Following the reviewers' suggestions that advised to apply a reorientation of the project, the strategy and the scheduling of the use cases and of the associated demonstrators was modified by the Steering Committee in a meeting held in Paris at the end of July, 2013. In this report we present motivations that led to the change and the current status that concerns the demonstrators.

1 Introduction

WP2/NA2 activity is based on the seismological community of practice—the main stakeholder of VERCE. The main activities of NA2 for this reporting period are provided in D-NA2.3 whereas here we restrict the discussion to the reasons that led to the change of scheduling of the present report which concerns the demonstrators of the use cases.

The development and the implementation of the use cases on the VERCE platform have been modified to take into account the results of the review meeting held in April 2013, in Paris. Specifically, the Steering Committee, SC, during a meeting which took place at Charles De Gaulle Airport on 30 July 2013 decided to give precedence to the HPC compute-intensive use case to the goal of providing by March 2014 a working and ready-to-be-tested beta-version. This use case entails the development of waveform simulation of earthquake events and the analysis of the misfit between the simulated waveforms and the observed waveforms at a given set of stations to evaluate the quality of the earth models. The second decision of the SC regarded the data-intensive use case which needs for the moment much attention on the following main barriers identified by the seismologists — the ingestion of the raw data sets with appropriate data structure organisation and formats, and the data management layer during the end-to-end workflow. This relevant reorientation of the use cases delayed the development of the demonstrators and the change of the work plan is discussed below.

2 Demonstrators and the modified work plan

The motivations that led to the change of the work plan can be summarised as follows.

1. The demonstrators presented at the review meeting in Paris in April 2013 were a proof-of-concept of the VERCE platform. At this stage they are not yet at a stage of a demonstrator for dissemination and public outreach and required additional work including an integration within the scientific gateway.

2. The reviewers indicated that we should give higher priority on the HPC use case for synthetic waveforms prediction. This is the demonstrator that has been set to first priority and the objective is to integrate that use case within a scientific gateway integrating SCI-BUS technology. One can elaborate there what should be the minimal requirement for such a demonstrator as well as the targeted audience which will be selected external seismological users by the end of next March. The external users will be selected within the EPOS and CIG community. It is noteworthy that this was well received by the community and, for example, the groups of J. Tromp and G. Nolet demonstrated much attention to the activities of VERCE. The main objective of this HPC use case demonstrator will be to demonstrate how VERCE can leverage data movement and job submission on PRACE HPC centres, i.e. LRZ and CINECA, together with providing a set of tool for preparing the input data and analysing the simulation results. Therefore, for the next period, the efforts will focus primarily on this demonstrator, and will lead to a training session in March 2014 open to those selected external users drawn from the EPOS and CIG communities above. The training session will in particular focus on the scientific gateway integration of this demonstrator and how external users can make use of it.

3. For the data-intensive application use case, according with the reviewers' comments, we have agreed to postpone the demonstrator focusing on improving the scalability and the data management together with fostering the contribution of the seismologists by allowing them to make use of Python-based workflow scripts. The priority here is to come up in March 2014 with an internal demonstrator that the seismologists of VERCE can evaluate.

In conclusion, we find that, although the development of the demonstrators has been delayed, there is, however, a clear development plan of the use cases with clear prioritisation. We expect that by March

2014 the HPC synthetic waveform one will be made available and tested by the community of reference and by the same time, it is expected the data-intensive demonstrator for internal use.