



Contribution ID: 15

Type: Oral presentation

Reproducible science at large scale within a continuous delivery pipeline: the BSC vision

Tuesday, 15 October 2019 12:20 (20 minutes)

Numerical models of the climate system are an essential pillar of modern climate research. Over the years, these models have become more and more complex and today's Earth System Models (ESMs) consist of several components of the climate system coupled together, and running on high performance computing (HPC) facilities. Moreover, climate experiments often entail running different instances of these models from different starting conditions and for an indeterminate number of steps. This workflow usually involves other tasks needed to perform a complete experiment, as data pre-processing, post-processing, transferring or archiving.

As a result, to reproduce a climate experiment is far from a trivial task which requires the orchestration of different methodologies and tools in order to guarantee, when not bit to bit, the statistical reproducibility of the research.

In this work we show the methodology and software tools employed to achieve science reproducibility in the Earth Sciences department of the Barcelona Supercomputing Center. Version control systems (VCS), test-oriented development, continuous integration with automatic and periodic tests, and a well-established reproducibility methodology, data repositories and data federation services, are all orchestrated by a fault-tolerant workflow management system.

Additionally, we show our experience in providing an operational service in the context of a research environment, with the set up of a highly fault-tolerant system. In this case the option of providing redundancy by using cloud services to execute computational models was considered and studied in comparison with the capabilities provided by HPC systems.

Primary authors: CASTRILLO, Miguel (BSC-CNS); Dr ACOSTA COBOS, Mario (BSC); SERRADELL MARONDA, Kim (Barcelona Supercomputing Center)

Presenter: CASTRILLO, Miguel (BSC-CNS)

Track Classification: Workshop: Building reproducible workflows for earth sciences