

# Programming Distributed Computing Platforms with COMPSs

Wednesday, 11 November 2015 16:00 (2 hours)

Distributed computing platforms like clusters, grids and clouds pose a challenge on application developers due to different issues such as distributed storage systems, complex middleware, geographic distributions. COMPSs [1] is a programming model which is able to exploit the inherent concurrency of sequential applications and execute them in a transparent manner to the application developer in distributed computing platforms.

This is achieved by annotating part of the codes as tasks, and building at execution a task-dependence graph based on the actual data consumed/produced by the tasks. The COMPSs runtime is able to schedule the tasks in the computing nodes and take into account facts like data locality and the different nature of the computing nodes in case of heterogeneous platforms. Additionally, recently COMPSs has been enhanced with the possibility of coordinating Web Services as part of the applications and extended on top of a big data storage architectures.

In the course, the syntax, programming methodology and an overview of the runtime internals will be given. The attendees will get a first lesson about programming with COMPSs that will enable them to start programming with this framework. The attendees will analyze several examples of COMPSs programming model compared with other programming models, such as Apache Spark, and also examples of porting libraries and codes to this framework. Different programming languages will be used including Java and Python whose adoption for scientific computing has been gaining momentum in the last years [2].

A hands-on with simple introductory exercises will be also performed. The participants will be able to develop simple COMPSs applications and to run them in the EGI Federated Cloud testbed. COMPSs is available in the EGI Cloud Marketplace as solution [3] for the integration of applications (use cases from BioVeL, LOFAR and EUBrazilCC communities) in the federated cloud environment providing scalability and elasticity features.

## Links, references, publications, etc.

[1] F. Lordan, E. Tejedor, J. Ejarque, R. Rafanell, J. Álvarez, F. Marozzo, D. Lezzi, R. Sirvent, D. Talia, R. M. Badia: ServiceSs: An Interoperable Programming Framework for the Cloud. *J. Grid Comput.* 12(1): 67-91 (2014).

[2] E. Tejedor, Y. Becerra, G. Alomar, A. Queralt, R. M. Badia, J. Torres, T. Cortes, J. Labarta, PyCOMPSs: Parallel Computational Workflows in Python, to be published in *International Journal of High Performance Computing Applications*.

[3] D. Lezzi, F. Lordan, R. Rafanell, R. M. Badia, "Execution of scientific workflows on federated multi-cloud infrastructures", *Euro-Par 2013: Parallel Processing Workshops Lecture Notes in Computer Science Volume 8374*, 2014, pp 136-145.

[www.comps.bsc.es](http://www.comps.bsc.es) with links to tutorial material, manuals and software that will be used in the course.

## Additional information

The tutorial will be part of BSC training program that is also linked with Prace PATC courses <http://www.bsc.es/marenostrum-support-services/hpc-trainings/prace-trainings>

These courses are officially credited by the UPC (Barcelona Tech University) Informatics Faculty (FIB). BSC also organizes COMPSs training for the Spanish Supercomputing Network (RES).

Prerequisites: programming skills in Java; basic knowledge of the EGI Federated Cloud testbed. The attendees should bring their laptop to perform the hand-on part of the tutorial.

**Primary author:** LEZZI, Daniele (Barcelona Supercomputing Center)

**Co-authors:** Mr DIAZ, Carlos (Barcelona Supercomputing Center); Mr RAMON-CORTES, Cristian (Barcelona Supercomputing Center); Mr LORDAN, Francesc (Barcelona Supercomputing Center); Dr CONEJERO, Javier

(Barcelona Supercomputing Center); Mr EJARQUE, Jorge (Barcelona Supercomputing Center); Dr SIRVENT, Raul (Barcelona Supercomputing Center); Dr BADIA, Rosa Maria (Barcelona Supercomputing Center)

**Presenter:** LEZZI, Daniele (Barcelona Supercomputing Center)

**Session Classification:** Tutorial: Programming Distributed Computing Platforms with COMPSs