Contribution ID: **82**   Type: **not specified**

Enhancement of the FutureGateway and workflows frameworks in order to support within INDIGO Platform the use cases provided by the INDIGO Communities

Wednesday, 6 April 2016 16:40 (20 minutes)

In Cloud computing, both the public and private sectors are already offering Cloud resources as IaaS (Infrastructure as a Service). However, there are numerous areas of interest to scientific communities where Cloud Computing uptake is currently lacking, especially at the PaaS (Platform as a Service) and SaaS (Software as a Service) levels.

In this context, INDIGO-DataCloud (INtegrating Distributed data Infrastructures for Global ExplOitation) [1], a project funded under the Horizon 2020 framework program of the European Union, aims at developing a data & computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over hybrid e-Infrastructures. This platform features contributions from leading European distributed resource providers, developers, and users from various Virtual Research Communities (VRCs). It is based on open source solutions addressing scientific challenges in the Grid, Cloud and HPC/local infrastructures and, in the case of Cloud platforms, providing PaaS and SaaS solutions.

SaaS solutions are exposed to end user through Science Gateways, mobile appliances, and APIs to be integrated in desktop applications. INDIGO adopts the Future Gateway (FG) [2] framework as both the presentation layer and the API server for the end user applications.

The FG is a standard-based solution that, by exploiting well consolidated standards like OCCI, SAGA, SAML, TOSCA, etc., is capable to target many distributed computing infrastructure, while providing a solution for mobile appliances as well. We will present the latest developments of the FutureGateway that is an evolution of the Catania Science Gateway[13]. We will demonstrate "live" two use cases selected by the project from the final users' perspective. They are making use of the Future Gateway framework, scientific workflows (like Kepler) and big data analytics tools (like Ophidia).

They are briefly explained in the following.

1) Climate change: the case study on Climate models intercomparison data analysis relates to the climate change domain and community (European Network for Earth System modelling - ENES [3]). It is directly connected to the Coupled Model Intercomparison Project (CMIP), one of the most internationally relevant and large climate experiment as well as to the Earth System Grid Federation (ESGF)[4][5] infrastructure in terms of existing eco-system and services. In the last three years, ESGF has been serving the Coupled Model Intercomparison Project Phase 5 (CMIP5 [6]) experiment, providing access to 2.5PB of data for the IPCC AR5 [7][8]. The test case focuses on a subset of this global data archive and proposes a common approach to perform three different scientific data analysis classes: (i) trend analysis, (ii) anomalies analysis, and (iii) climate change signal analysis. The first one will be specifically addressed by the demo. The test case demonstrates the INDIGO capabilities in terms of software framework deployed on heterogeneous infrastructures (e.g., HPC clusters and cloud environments), as well as workflow support to run distributed, parallel data analyses. While in this use case general-purpose WfMSs (in this case Kepler WfMS[9]) are exploited to orchestrate multi-site tasks, the Ophidia framework [10][11] is adopted at the single-site level to run scientific data analytics workflows consisting of tens/hundreds of data processing, analysis, and visualization operators. The demonstration will highlight: (i) the interoperability with the already existing community-based software ecosystem and infrastructure (IS-ENES/ESFG); (ii) the adoption of workflow management system solutions (both coarse and fine grained) for large-scale climate data analysis (e.g. Ophidia, Kepler); (iii) the exploitation of Cloud technologies/solutions from the INDIGO PaaS offering easy-to-deploy, flexible, isolated and dynamic big data analysis solutions; and (iv) the provisioning of interfaces, toolkits and libraries to develop high-level interfaces/applications integrated in a Science Gateway. With regard to the last point, the demo will show how the results of the experiments will be easily made available to the end user for inspection, download, and visualization. To this end, the user interface will provide specific/advanced support for data analytics and visualization.

2) Molecular Dynamics of proteins: the three-dimensional (3D) structure of biological macromolecules consists of a set of (x,y,z) coordinates for each atom of the molecule under investigation. The INSTRUCT ESFRI [12] provides access to high specification, specialist equipment for the experimental determination of such coordinates. However, the 3D structure of any molecule is not completely rigid, but fluctuates over time due to

the kinetic energy available at room temperature. Such flexibility is often directly relevant to the physiological function performed by proteins and nucleic acids in the cell. Although there are experiments that can provide information on the extent and time scales of macromolecular motions, computer simulation (Molecular Dynamics, MD) is the only technique that provides a full atomistic view of motions throughout all regions of the macromolecule. The present demo will highlight the use of the exploitation of Cloud technologies/solutions from the INDIGO PaaS to perform MD simulations using protocolized methods in VMs and the use of web interfaces to set up and analyze such simulations.

References
[1] https://www.indigo-datacloud.eu/
[2] https://www.indigo-datacloud.eu/documents/software-architecture-and-work-plan-wp6-d61
[3] European Network for Earth System modelling - https://verc.enes.org/community/about-enes
[4] Earth System Grid Federation - http://esgf.llnl.gov
[5] Luca Cinquini, Daniel J. Crichton, Chris Mattmann, John Harney, Galen M. Shipman, Feiyi Wang, Rachana Ananthakrishnan, Neill Miller, Sebastian Denvil, Mark Morgan, Zed Pobre, Gavin M. Bell, Charles M. Doutriaux, Robert S. Drach, Dean N. Williams, Philip Kershaw, Stephen Pascoe, Estanislao Gonzalez, Sandro Fiore, Roland Schweitzer: The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. Future Generation Computer Systems 36: 400-417 (2014).
[6] Coupled Model Intercomparison Project Phase 5 (CMIP5) - http://cmip-pcmdi.llnl.gov/cmip5/
[7] Intergovernmental Panel on Climate Change –http://www.ipcc.ch
[8] IPCC Fifth Assessment Report - https://www.ipcc.ch/report/ar5/
[9] Marcin Płóciennik, Tomasz Żok, Ilkay Altintas, Jianwu Wang, Daniel Crawl, David Abramson, Frederic Imbeaux, Bernard Guillerminet, Marcos Lopez-Caniego, Isabel Campos Plasencia, Wojciech Pych, Pawel Ciecieląg, Bartek Palak, Michał Owsiak, and Yann Frauel. 2013. Approaches to Distributed Execution of Scientific Workflows in Kepler. Fundam. Inf. 128, 3 (July 2013), 281-302.
[10] Sandro Fiore, Alessandro D'Anca, Cosimo Palazzo, Ian T. Foster, Dean N. Williams, Giovanni Aloisio: Ophidia: Toward Big Data Analytics for eScience. ICCS 2013: 2376-2385
[11] Sandro Fiore, Cosimo Palazzo, Alessandro D'Anca, Ian T. Foster, Dean N. Williams, Giovanni Aloisio: A big data analytics framework for scientific data management. BigData Conference 2013: 1-8
[12] https://www.structuralbiology.eu/
[13] http://www.catania-science-gateways.it/

Presenters: GIACHETTI, Andrea (CIRMMP); ROSATO, Antonio (CIRMMP); GIORGIO, Emidio (INFN); PLOCIENNIK, Marcin (ICBP); FARGETTA, Marco (INFN); OWSIAK, Michal; BRUNO, Riccardo (INFN); BARBERA, Roberto (University of Catania and INFN); FIORE, Sandro (SPACI); ZOK, Tomasz (ICBP)

Session Classification: Federated cloud PaaS and SaaS