

Next generation Science Gateways in the context of the INDIGO project: a pilot case on large scale climate change data analytics

Thursday, 12 November 2015 16:10 (20 minutes)

The INDIGO project 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). INDIGO aims to develop tools and platforms 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 that are currently lacking for e-Science. INDIGO will also develop a flexible and modular presentation layer connected to the underlying IaaS and PaaS frameworks, thus allowing innovative user experiences including web/desktop applications and mobile appliances. INDIGO covers complementary aspects such as VRC support, software lifecycle management and developers support, virtualized resource provisioning (IaaS), implementation of a PaaS layer and, on a top level, provisioning of Science Gateways, mobile appliances, and APIs to enable a SaaS layer. INDIGO adopts the Catania Science Gateway framework (CSGF) as presentation layer for the end users. The CSGF is a standard-based solution that, by exploiting well consolidated standards like OCCI, SAGA, SAML, etc., is capable to target any distributed computing infrastructure, while providing a solution for mobile appliances as well. In the context of INDIGO, the CSGF will be completely re-engineered in order to include additional standards, such as CDMI and TOSCA, and to be exposed as a set of APIs.

This paper presents an early use case examined by the project from the final users perspective, therefore interfacing INDIGO targeted resources through a preliminary web interface, provided by a Science Gateway, which hides the complexities of the underlying services/systems.

This use case relates to the climate change domain and community (European Network for Earth System modelling - ENES) and tackles large scale data analytics requirements related to the CMIP5 experiment, and more specifically to anomalies analysis, trend analysis and climate change signal analysis. It 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 general-purpose WfMSs (e.g., Kepler, Taverna) are exploited in this use case to orchestrate multi-site tasks, the Ophidia framework 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 contribution will highlight: (i) the interoperability with the already existing community-based software eco-system and infrastructure (IS-ENES/ESGF); (ii) the adoption of workflow management system solutions (both coarse and fine grained) for large-scale climate data analysis; (iii) the exploitation of Cloud technologies 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. The presentation will also include a discussion on how INDIGO services have been designed to fulfil the requirements of many diverse VRCs.

Primary authors: SALOMONI, Davide (INFN CNAF); GIORGIO, Emidio (INFN Catania); DONVITO, Giacinto (INFN Bari); ALOISIO, Giovanni (University of Salento and CMCC); PLOCIENNIK, Marcin (PSNC); FARGETTA, Marco (INFN Catania); BRUNO, Riccardo (INFN Catania); BARBERA, Roberto (University of Catania and INFN); FIORE, Sandro (CMCC)

Presenter: BARBERA, Roberto (University of Catania and INFN)

Session Classification: Virtual Research Environments