Contribution ID: **76** Type: **Presentation**

Managing Virtual Research Environments in Hybrid Data Infrastructures

Tuesday, 18 September 2012 11:22 (22 minutes)

Description of the work

Science is increasingly global, multidisciplinary and networked. It needs access to a large amount of datasets that come in all forms and shapes from huge international experiments to cross-laboratory, single laboratory, or even from a multitude of individual observations. The exploitation of datasets originally maintained by several organizations usually spread worldwide represents the new challenging requirement. It is not even possible to think to a future where few selected standards, best practices, and policies will be widely adopted by the science ecosystem. Heterogeneity will likely continue to exist even in the presence of an increasing multidisciplinary approach to science.
Hybrid Data Infrastructures (HDI) born to deal with such heterogeneity. They integrate several technologies for data management, access and analysis while providing transparent access to heterogeneous computational and storage platforms. Moreover, HDI preserving management capabilities such as monitoring, accounting, and secure access become a credible approach in this new challenging scenario.
The gCube software system implements the HDI approach. It offers a data-management-capability-delivery model in which computing, storage, data and software are made accessible by the infrastructure and are exploited by users using a thin client (namely a web browser), through dedicated on-demand Virtual Research Environments (VRE).
VREs are declaratively and dynamically build while abstracting on the location, provenance and interfaces of the resources. gCube technology implements a user friendly Software as a Service framework where the data, the application services, and the storage and computing resources needed by a scientist are automatically aggregated and made available through a web based interface. The aggregated resources are also monitored to guarantee the VRE service while guaranteeing secure and controlled access.

Link for further information

D4Science : www.d4science.org
gCube web site: www.gcube-system.org
gCube documentation: gcube.wiki.gcube-system.org/gcube

Wider impact of this work

In this emerging new science, the HDI maintainer aggregates computational and storage resources from a variety of providers including commercial Cloud providers, while resource consumers buy them only for the time needed for their exploitation and use them to build their virtual environments.
Thanks to D4Science HDI and the gCube technology for VRE creation, for example a newly Biodiversity VRE:
- aggregates datasets from the Ocean Biogegraphic Information System, the Global Biodiversity Information Facility, the Ocean Monitoring and Forecast initiative, the NCBI, the World Register of Marine Species, and others
- offers new and unprecedented computing capabilities obtained accessing transparently to resources provided by D4Science contributors and by Windows Azure Cloud computing platform
- integrates technologies for managing ecological niche modelling, time series harmonization, statistical data analysis with R, and data mining with WEKA.

Printable Summary

The global and networked needs of the science produce and need to exploit huge quantities and varieties of data. To confirm this impression, a recent study, promoted by The Royal Society of London, highlighted how science is becoming increasingly global, multidisciplinary and networked.
Singleton technological platforms are no longer able to address the data and processing requirements of the emerging data-intensive science characterized by a predominant data distribution and by evolving user communities. A novel approach, the Hybrid Data Infrastructure (HDI), integrates several technologies, including Grid and Cloud, and offers the necessary management capabilities required by geographically disperse user communities and data centers.
If HDI provides the data processing and analysis computing capabilities, Virtual Research Environments allow to manage the elastic and secure aggregation of users in focused communities living for the time needed to achieve their objective.

Primary authors: MANZI, Andrea (CERN); Dr CANDELA, Leonardo (CNR - ISTI); PAGANO, Pasquale (CNR - ISTI)

Presenter: PAGANO, Pasquale (CNR - ISTI)

Session Classification: Research Infrastructures

Track Classification: Virtual Research Environments (Gergely Sipos: track leader)