Contribution ID: **64** Type: **Presentation**

D4Science Infrastructure: a novel open approach to distributed software management based on Maven

Thursday, 20 September 2012 11:40 (20 minutes)

Description of the work

gCube, the D4Science infrastructure powering technology, is a JAVA software framework featuring the declarative and interactive creation of transient Virtual Research Environments by aggregating and deploying on-demand content resources and application services. It has been designed to exploit the peculiarities of modern application servers (JAVA WS-core, Tomcat, Jboss), which ensure Web-services/web-applications "hot"deployment, and to bridge existing private and public cloud systems. Since its inception, the focus of the framework has been on the optimal resource allocation and management of software packaged following the gCube policies. This was indeed a weak point of the system imposing custom packaging rules and limiting de-facto the software potentially deployable on the infrastructure. To overcome it, the gCube enabling layer recently took an extended approach for dealing with software available as artifact in an approved Maven Repository. This also had the great advance of introducing Maven as official distribution repository for gCube software. The approach firstly comprised the exploitation of Maven within the build and integration process. Then, it required the development of a new service, the Software Gateway, which acts as a gateway over a cluster of Maven Repositories granting access to the stored information and software for deployment purposes. And finally, all the rest of the gCube services managing software were adapted to exploit the new deployment model, not based on the previous policies.We will report on this new open approach along with the extensions applied to the gCube resource model to cope with maven artifacts. To give a complete picture of the impact on the development, we will cover major aspects of the gCube build process, the "mavenization"of the legacy gCube artefacts and the activities performed at integration side to deal with Maven. Lastly, a look ahead to future activities to support deployment of non-JAVA based artifacts.

Link for further information

http://www.gcube-system.org/
https://gcube.wiki.gcube-system.org/gcube/index.php/Data_e-Infrastructure_Management_Facilities

Wider impact of this work

Simplified management and deployment of software coming from third-party repositories is one of the main goals in a growing and open environment as the D4Science infrastructure. Communities coming from different EU projects, with different requirements in term of software management are today exploiting the infrastructure. The introduction of a de-facto standard as Maven orthogonal to many aspects of software management has been a major step forward towards interoperability and sustainability. From one side we enable users' applications to be deployed on the infrastructure at nearly "zero" cost. From the other side, integrating Maven grants access to gCube software to other Maven-systems and opens the possibility to adopt Maven at different levels. This is a crucial point to evaluate as the end of EMI project's end is approaching and the future of its building facilities, greatly exploited by gCube since many years, is yet unsettled.

Printable Summary

The D4Science infrastructure is a Hybrid Data Infrastructure (HDI) deployed and maintained throughout three EU projects (DILIGENT, D4Science I and II) and actually supporting two EU projects (iMarine and EU-BrazilOpenBio). A Hybrid Data Infrastructure is an innovative approach based on the integration of several technologies, including Grid and private/public Clouds, to provide an elastic access to and usage of data and

data-management capabilities. Equipped with services supporting the creation of Virtual Research Environments, it creates dynamic and distributed applications tailored to serve a specific need whose constituents are acquired by the HDI. We report on a major extension performed on the infrastructure: the deployment and activation of Maven artifacts. This is a key aspect for the sustainability of the infrastructure promoting the transparent exploitation of third-party applications in a Virtual Research Environment.

Primary authors: MANZI, Andrea (CERN); Dr SIMEONI, Fabio (FAO); Mr GIAMMATTEO, Gabriele (ENG); Mr SIMI, Manuele (ISTI-CNR)

Co-author: PAGANO, Pasquale (CNR - ISTI)

Presenter: MANZI, Andrea (CERN)

Session Classification: Virtualised Resources

Track Classification: Virtualised Resources: challenges and opportunities (Michel Drescher: track leader)