gCube: A Service-Oriented Application Framework on the Grid

by Leonardo Candela, Donatella Castelli and Pasquale Pagano

gCube is a new service-oriented application framework that supports the on-demand sharing of resources for computation, content and application services. gCube enables the realization of e-infrastructures that support the notion of Virtual Research Environments (VREs), ie collaborative digital environments through which scientists, addressing common research challenges, exchange information and produce new knowledge. gCube is currently used to govern the e-infrastructure set up by the European integrated project DILIGENT (A Digital Library Infrastructure on Grid-Enabled Technology).

The long-term journey towards the e-science vision demands e-infrastructures that allow scientists to collaborate on common research challenges. Such infrastructures provide seamless access to necessary resources regardless of their physical location. These shared resources can be of very different natures and vary across application domains. Usually they include content resources, application services that manipulate these content resources to produce new knowledge, and computational resources, which physically store the content and support the processing of the services. Many e-infrastructures already exist, at different levels of maturity, and support the sharing of and transparent access to resources of a single type, eg SURFNet (information), GriPhyN (services), EGEE (computing and storage). However, they are still too primitive to support a feasible realization of VREs. The DILIGENT infrastructure, released recently by the homonymous project, will overcome this limitation by supporting in a single common framework the sharing of all three types of resource.

The core technology supporting such e-infrastructure is a service-oriented application framework named gCube. gCube enables scientists to declaratively and dynamically build transient VREs by aggregating and deploying on-demand content resources, application services and computing resources. It also monitors the shared resources during the lifetime of the VRE, guaranteeing their optimal allocation and exploitation. Finally, it provides mechanisms to easily create dedicated Web portals through which scientists can access their content and services.

From the technological point of view, gCube provides: (i) runtime and design frameworks for the development of services that can be outsourced to a Grid-enabled infrastructure; (ii) a service-oriented Grid middleware for exploiting the Grid and hosting Web Services on it; (iii) a set of application services for distributed information management and retrieval of structured and unstructured data.
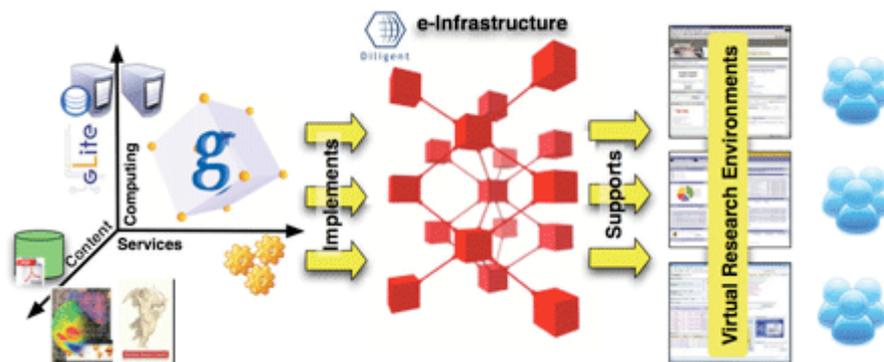

Figure 1: gCube in action.

Runtime frameworks are distinguished workflows that are partially pre-defined within the system; they include Grid-enabled services and application services, where the former coordinate in a pure distributed way the action of the latter, while relying on a high-level characterization of their semantics. Design frameworks consist of patterned blueprints, software libraries and partial implementations of state-of-the-art application functionality, which can be configured, extended and instantiated into bespoke application Grid services.

The service-oriented Grid middleware provides all the required capabilities necessary to manage Grid infrastructures. It eliminates manual service deployment overheads, guarantees optimal placement of services within the infrastructure and opens unique opportunities for outsourcing state-of-the-art implementations. Rather than interfacing with the infrastructure, the software which implements the application services is literally handed over to it, so as to be transparently deployed across its constituent nodes according to functional constraints and quality-of-service (QoS) requirements. By integrating the gLite system released by the Enabling Grid for E-sciencE (EGEE) project for batch processing and management of unstructured data, gCube also allows the large computing and storage facilities provided by the EGEE infrastructure to be properly exploited. With over 20,000 CPUs and 5 million Gigabytes of storage, EGEE is the largest operational Grid infrastructure ever built.

gCube application services offer a full platform for distributed hosting, management and retrieval of data and information, and a framework for extending state-of-the-art indexing, selection, fusion, extraction, description, annotation, transformation and presentation of content. In particular, gCube is equipped with services for manipulating information objects, importing external objects, managing their metadata in multiple formats, securing the information objects to prevent unauthorized access, and transparently managing replication and partition on the Grid.

In order to promote interoperability, gCube services are implemented in accordance with second-generation Web service standards, most noticeably SOAP, BPEL, WSRF, WS-Notification, WS-Security, WS-Addressing, and JSR168 Portal and Portlets specifications.

gCube is the result of the collaborative efforts of more than 48 researchers and developers in twelve different academic and industrial research centres: Institute of Information Science and Technologies ISTI-CNR (IT), University of Athens (GR), University of Basel (CH), Engineering Ingegneria Informatica SpA (IT), University of Strathclyde (UK), FAST Search & Transfer (NO), CERN European Organisation for Nuclear Research (CH), 4D SOFT Software Development Ltd (HU), European Space Agency ESA (FR), Scuola Normale Superiore (IT), and RAI Radio Televisione Italiana (IT), and ERCIM.

Information about the project and detailed information about the software, currently available in its beta version, can be found at the gCube Web site.

Links:
gCube: http://www.gcube-system.org/
DILIGENT: http://www.diligentproject.org/
EGEE: http://www.eu-egee.org/

Please contact:
Donatella Castelli, Pasquale Pagano, ISTI-CNR, Italy
E-mail: donatella.castelli@isti.cnr.it, pasquale.pagano@isti.cnr.it