The EGI pan-European Federation of Clouds

Matteo Turilli

Oxford e-Research Centre, University of Oxford, United Kingdom.
emails: matteo.turilli@oerc.ox.ac.uk

Keywords: Infrastructure as a Service, clouds federation, EGI

The European Grid Infrastructure (EGI) is the result of a coordinated effort that, over the last decade, has supported multi-disciplinary science by federating national resource providers into a pan-European production infrastructure. Currently, EGI fosters an ecosystem of funding agencies, coordinating bodies, technology providers and integrators, resource providers, and more than three hundred resource centres. Together, they serve twenty-one thousand researchers across fifteen disciplines carrying out a total of 1.2 million computing jobs a day.

Until now, EGI has championed grid technologies to federate resources across Europe and the world but the number of resource centres evaluating and deploying private clouds to support their institutional researchers is growing at a fast pace. Cloud computing is attractive for both providers and users. Providers can leverage virtualisation and isolate resource management from the support of user-facing applications and services. User communities find in cloud computing an efficient model of resource distribution that, by offering flexibility, accessibility and dynamic provisioning, suits well the requirements of digital science.

As done already with grid computing, cloud resources have to be federated to support user communities that need to share resources across multiple research institutions. Unfortunately, federating cloud resources is not a straightforward process. Still in an early stage of development, cloud computing is implemented by a heterogeneous set of technologies that involve different hypervisors and multiple interfaces to manage computing, storage and network resources. As a consequence, the European cloud landscape is fragmented, with resource providers deploying different types of cloud platforms while exposing proprietary interfaces and implementing different sets of capabilities.

In 2011, the EGI-InSPIRE project established a Federated Cloud Task Force and transformed it into an official Task in 2012. The goal is to facilitate the setup of a pan-European federated cloud based on the resources of National Grid Infrastructures.

The talk describes the activities and deliverables of the Task Force focusing on the model that has been chosen to deploy an EGI pan-European clouds federation test bed. The talk is divided into three parts. First, the mandate and membership of the Task force are illustrated. The objectives of the EGI Federated Cloud Task Force are:

- produce a blueprint for EGI resource providers that wish to federate and share their virtualised environments as part of EGI;
- deploy a test bed to evaluate the integration of virtualised resources within EGI;
- investigate and catalogue the requirements of user communities;
- provide feedback to relevant technology providers on their implementations;
- identify and work with user communities willing to be early adopters of the federation test bed;
- identify issues that need to be addressed by other areas of EGI.

The Task Force has been active since September 2011. After an initial engagement phase, twenty-five institutions from sixteen European countries have been actively participating to

the Task Force operations. Of these, eighteen institutions are providing cloud resources to set up a federation test bed and seven technology providers are contributing the required centralised services and software integration effort. Until now, eight user communities have joined the Task Force offering requirements, use cases and prototype services for research. The Task Force endorses an inclusive and open approach for its operations and four liaisons have been set up with other initiatives operating on cloud computing and clouds federation.

In the second part of the talk, several use cases are reviewed to show how user communities intend to benefit from cloud computing and what kind of capabilities a federation of clouds should offer. The use cases include the humanities in the form of the British National Corpus (through BNCweb), which contains a dataset of 100 million English words and PeachNote, with a database of 1.6 million musical scores. Both of these activities require intensive processing in order to support researchers in these areas and can leverage a dynamic and distributed model of resource provisioning. On the scientific side, structural biologists want to scale out virtualised services across a wide range of resources for both teaching and research needs. Astronomers need to capitalise on the flexibility offered by the cloud model, with large data sets of astronomical data that have to be accessed and interrogated by various applications. Finally, the Taskforce is working with the science gateways project SCI-BUS, in order to provide seamless access to a broader range of federated resources.

The third part of the talk reviews different models of federation and offers a detailed analysis of the one chosen by EGI. The test bed deployed by the Task Force embodies a light-weight federation model based on standards and centralised services. Resource providers are not requested to deploy any specific cloud management platform in order to contribute their resources to the federation. Moreover, resource providers remain in charge of their local usage policies, both in term of quantity of resources to be federated and security policies. The federation layer is made of two management interfaces and four services:

- an Open Cloud Computing Interface (OCCI) for the management of VMs, storage volumes and network resources;
- a Cloud Data Management Interface (CDMI) client/server system used to move data to and from the instantiated VMs and data repositories;
- an information system based on an extended version of the GLUE2 schema, a top-BDII server, and a LDAP server installed in each federated Resource Provider;
- a monitoring system based on Nagios with a set of probes developed specifically to monitor the status of the federated services and management interfaces;
- a metadata publishing service called Marketplace used to advertise the images from which Virtual Machines (VMs) can be instantiated in the test bed;
- an accounting infrastructure based on an extended OGF Usage Record schema and implemented on an APEL system.

End-users are provided with an OCCI compliant, command line client that allows for using resources on multiple providers with a high degree of transparency and centralisation.

The talk closes with a discussion of the strategy for the integration of cloud resources into the existing EGI production infrastructure. Particular attention will be given to resource brokering and how cloud provisioning compares to other ways to share and distribute computing and data resources to the research communities.