



Contribution ID: 63

Type: **Oral**

OpenMOLE: a grid enabled workflow platform

Thursday, 15 April 2010 09:20 (20 minutes)

OpenMOLE is a free and open source workflow engine providing distributed computing facilities, especially suited for scientific research in complex systems. Third-party software packages can be embedded in a workflow that automatically transfers and processes their input and output (files and data). Embedded software packages are called “tasks”. Any task within a workflow can be either executed locally or delegated to distributed computing environments, including the EGEE grid.

Detailed analysis

The main purpose of OpenMOLE is to provide a high-level workflow platform for designing scientific experiments. OpenMOLE decouples the scientific business logic from the resources used to execute it. It enables the definition of scientific workflows and the delegation of workflow tasks to high-performance computing environments in a declarative way.

In a first phase, we have conceived workflow tasks in such a way that the platform is able to migrate them on demand to another execution environment than the end-user local personal computer (PC). Thus we have defined what is an execution context and the consumed resources of a task. The implementation of these concepts allows remote execution of a given task.

We also faced the challenge of establishing direct links between a user PC and external distributed computing environments. Although OpenMOLE does not assume that the user PC owns a public IP address, no particular third-party server (i.e., other than the user PC and the execution environment servers) is ever required. Furthermore, no preliminary installation step on a remote environment is required: OpenMOLE transports everything it needs to be run remotely.

Conclusions and Future Work

The core functionalities of the workflow engine have been implemented. The greatest part of our effort will be spent next on the development of a user-friendly GUI to take full advantage of OpenMOLE. A component is also under construction to keep track of experiments carried out in OpenMOLE and allow collaborative design of workflows among scientific communities. In parallel to platform development, more specific applications and scientific methods will be tested to ensure the broadest coverage of the diverse and interdisciplinary scientific domains composing complex systems.

Impact

Taking advantage of grid computing remains a tricky issue for the non-expert user. Software and hardware are heterogeneous, bad workload management decisions happen and, generally, failure rate is higher than other computing environments. In this context, a certain amount of technical and methodological knowledge must be acquired before making efficient use of grid computing. Fortunately, many types of applications and methods contain inherent parallel aspects. For these problems, we claim that a platform can completely hide the intricacies of execution environments to the user.

Software tools such as g-Eclipse, JSAGA or Ganga propose a high-level object layer to abstract the execution environment. Yet, they only partially hide the technical details and overall heterogeneity. OpenMOLE goes further and hides the whole distributed environment of the business layer.

Other software tools such as Taverna offer similar features. The OpenMOLE project, however, follows a different approach in which everything runs on the user desktop by default and no third-party server is ever called. Tasks are delegated on demand to distributed execution environments by the user.

Keywords

Workflow, generic platform, model exploration

URL for further information

<http://www.openmole.org/>

Primary author: Dr REUILLON, Romain (ISCPiF)

Co-authors: Mr LECLAIRE, Mathieu (ISCPiF); Mr DUMOULIN, Nicolas (CEMAGREF)

Presenter: Dr REUILLON, Romain (ISCPiF)

Session Classification: Workflow Management

Track Classification: End-user environments, scientific gateways and portal technologies