Contribution ID: **119**   Type: **Presentations**

An unattended, fault-tolerant approach for the execution of distributed applications

Tuesday, 9 April 2013 11:00 (20 minutes)

Impact

DistributedToolbox aims to help developers to speed up their application porting process, encapsulating the infrastructure-dependent operations away from the applications and allowing non expert users to create Grid-enabled applications as fast as possible.

With the employment of this proposal, all the possible problems related to the execution of tasks on dynamic environments are automatically managed by a devoted tool. The distributed infrastructure becomes then transparent to the developer.

Until now, five different applications have been ported and executed both on local clusters and Grid infrastructures. More than 75.000 tasks have been executed so far in production Grids belonging to two VOs without a single failure, demonstrating the stability of this proposal.

URL

http://www.ciemat.es/

Summary

In this work, the authors present a set of tools, DistributedToolbox to overcome the problem of executing distributed applications on dynamic environments. By employing an extremely simple interface to specify the characteristics and requirements of the tasks to be executed on the distributed infrastructure, the code developers can easily build distributed and portable applications.

The defined tasks can be executed either on local clusters (PBS and SGE out of the box, other alternatives are easy to implement) or on Grid Infrastructures on a completely unattended way. Within this approach, a task is considered to be executed when the desired output files are provided, ensuring that the distributed application will receive the required partial results.

Description

Currently, developers of distributed applications have to be aware of the details of the infrastructure where the distributed parts of their applications will run on. In the case of dynamic environments such as Grid infrastructures, fault detection and recovery must also be implemented. The authors of this work consider that this should not be a responsibility of the developer, so a toolbox called DitributedToolbox has been created to overcome this issue.

DistributedToolbox incorporated a small API so the tasks to be remotely executed can be defined, together with tools to execute them on different plattforms. These currently includes local clusters and Grid infrastructures. Because of the toolbox design, adding new computational plattforms is extremely easy.

The core of this work is GridController, a newly created tool for an unattended execution of tasks on Grid infrastructures. It is designed focusing on reliability, ensuring that after the user has specified a task to execute, the desired output files will be returned.

Based on GridWay metascheduler, GridController is able to detect any problem during the task execution such as problems on the remote resource, local site, authorization issues, data transfer or middleware failures, overcome them and execute the desired tasks. A small replication factor is employed to minimize the influence of slow sites on the execution time.

Primary author: Dr RODRÍGUEZ-PASCUAL, Manuel Aurelio (CIEMAT)

Co-authors: Mr RUBIO-MONTERO, Antonio Juan (CIEMAT); Dr MAYO GARCÍA, Rafael (CIEMAT)

Presenter: Dr RODRÍGUEZ-PASCUAL, Manuel Aurelio (CIEMAT)

Session Classification: Community Platforms

Track Classification: Community Platforms (Track Lead: P Solagna and M Drescher)

Primary author: Dr RODRÍGUEZ-PASCUAL, Manuel Aurelio (CIEMAT)

Co-authors: Mr RUBIO-MONTERO, Antonio Juan (CIEMAT); Dr MAYO GARCÍA, Rafael (CIEMAT)

Presenter: Dr RODRÍGUEZ-PASCUAL, Manuel Aurelio (CIEMAT)

Track Classification: Community Platforms (Track Lead: P Solagna and M Drescher)