



Contribution ID: 66

Type: **not specified**

GPU computing in EGI environment using a cloud approach

Monday, 11 April 2011 17:00 (30 minutes)

Overview

Recently GPU computing, namely the possibility to use the vector processors of graphics card as computational general purpose units of High Performance Computing environments, has generated considerable interest in the scientific community. Some communities in European Grid Initiative (EGI) are reshaping their applications to exploit this new programming paradigm. Each EGI community, called Virtual Organization (VO), often requires specific environments, making necessary for each grid site to enable an efficient system to fulfill VO's software requirements.

Cloud Computing and more generally the opportunity to transparently use computational resources, together with the consolidation of virtualization technologies, allows to provide to the end users the required environment for their activities.

The present work is aimed to provide for each VO a on-demand GPU environment (GPU framework, Operating System and libraries) and makes it accessible via the EGI infrastructure using the Cloud

Impact

The GPU computing is growing-up in EGI communities starting from Computational Chemistry (COMPCHEM VO) to Theoretical Physics (THEOPHYS VO) as well as from the needs of other communities.

The main purpose of the present work is to dynamically provide a ready to use GPU environment for the communities using the EGI infrastructure to share GPU resources, since a single GPU environment does not satisfy the different requirements of this communities (such as operating systems, compilers and scientific libraries). For this reason, the developed system provides dynamical environments with the aim to optimize GPU resources usage.

Contextually, the Cloud Computing opportunity allows to take into account the GPUs as a Service (IaaS). From a Cloud point of view, the project carries out a feasibility study to understand how the next evolution of the Grid Computing to Cloud Computing, or better, how the switch from Batch Model to Service Model could be done.

The approach adopted in this system is not focused only on GPU computing and it can be easily extended to other special hardware devices and, in general, to other environments (as occurs in the IaaS Model).

Description of the work

In the present work the possibility of enabling the convenient usage of GPUs devices for VO users, exploiting the capabilities of the EGI infrastructure, and the emerging paradigm of Cloud Computing, has been explored. A strategy to provide on-demand execution environments has been proposed through the joint usage of traditional and widespread gLite components and the popular standard EC2 web-service APIs.

An entire job flow that enables the Local Resource Management System (LRMS) to discriminate the GPU resources requests, through Glue Schema parameters, has been defined in order to allocate, in a dynamic fashion, the required resources on a Cloud-like infrastructure either public, private or hybrid.

To achieve this goal, part of the work has been devoted to the virtualization of the physical GPU resources in order to make them available in a Infrastructure as a Service (IaaS) private Cloud.

To this end a centralized mechanism, responsible to listen for events generated by the LRMS like job scheduling and termination, has been implemented to keep track of each request.

These events are then used to carry out the required actions as follows: once a job is received and identified as a GPU usage request, is treated as an event that triggers the allocation of virtualized resources according to simple leasing rules. In a similar way the termination of jobs are notified to a daemon that releases the execution environment.

In order to develop and test the whole infrastructure, a fully working test bed has been built with the adoption of the Eucalyptus software system to implement a private Cloud over the cluster.

We also addressed the need of the creation of Virtual Machine Images to match the requirements of the execution of GPU-dependent jobs, such as CUDA, OpenCL libraries and gLite middleware.

Conclusions

In this work we presented a system to provide to EGI users a specific on-demand GPU environment to transparently execute jobs on GPU devices.

The proposed system uses a Cloud approach, based on EC2 Compliant Clouds, in order to control the specific GPU-enabled VO's environment from EGI middleware interfaces.

The system is currently in its testing phase at UNI-PERUGIA Grid Site and supports COMPCHEM VO. In this phase the LRMS is used to join the EGI infrastructure and the UNI-PERUGIA private Cloud. We are planning to decouple the on-the-Cloud allocation mechanism from the LRMS and place it at the Computing Element Level using for example the CREAM Architecture capabilities.

This will allow a fine grained control over the Virtual Instances and the Accounting.

Primary authors: Dr COSTANTINI, Alessandro (UNIPG); Dr VELLA, Flavio (UNIPG); Prof. GERVASI, Osvaldo (UNIPG); Dr CEFALA', Riccardo (UNIPG)

Presenter: Prof. GERVASI, Osvaldo (UNIPG)

Session Classification: Virtualisation and Cloud Computing