GaaS: toward a more "elastic" and sustainable grid environment

**Dr. BARONE, Giovanni Battista (University of Naples Federico II), Dr. BIFULCO, Roberto (University of Naples Federico II),
Dr. BOCCIA, Vania (INFN Naples) , Dr. BOTTALICO, Davide (University of Naples Federico II), Dr. CARRACCIUOLO, Luisa (CNR)**

The scientific community developed the Grid Computing paradigm to enable the sharing of huge amount of resources through a well-defined distributed infrastructure model, in order to solve large scale problems in a collaborative manner. The Grid Computing resources aggregation model is rather "static": a group of organizations set up several Grid management services and computing resources in a layered structure. Users belonging to the organizations forming the Grid can retrieve information on resources (e.g., their number, status, configuration, etc.) and access them, but can neither change the topology of the grid (e.g., by increasing the number of resources) or manage resources configuration and composition. It would be desirable to have a more "elastic" infrastructure in which users can ask for resources on-demand, to suit their needs in terms of resources type and configuration (i.e compilers, scientific libraries, problem solving environments, etc.). Given the flexibility in resources management through the Cloud Computing paradigm, it seems a promising approach to provide flexible Grid Computing infrastructures through the combination of the Grid and Cloud paradigms We describe our experience in designing and implementing a solution that, by a "Grid over Cloud" approach, creates more flexible Grid infrastructures, by exploiting IaaS-provided resources, in a novel way resembling the PaaS paradigm. We call our solution Grid as a Service (GaaS).

The aim

to have a **more *"elastic"* infrastructure** in which users can ask for **resources *on-demand***, to suit their needs in terms of resources type and configuration (i.e compilers, scientific libraries, problem solving environments, etc.).

The problem

The typical Grid usage model does not allow the user in customizing the grid environment: users cannot change the Grid infrastructure that runs their experiment. In particular, an user cannot create a new Grid site nor add an existing site for his VO.

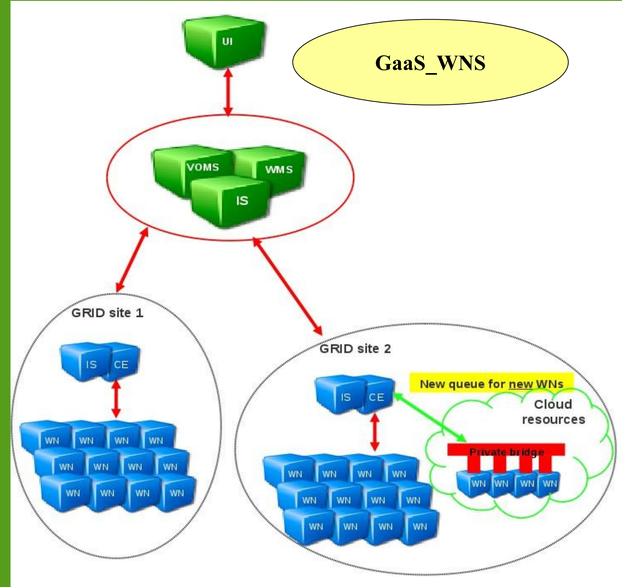
Grid sites are static and cannot be customized by users, hence, it is not possible to add new worker nodes to a site to extend its capabilities, and it is also not allowed to organize resources into customized queues to shape them in accordance to the computation needs, finally the configuration of the worker nodes cannot be changed.

Our solution

The GaaS model combines the advantage of providing users with an usage model that is familiar to the traditional Grid, with the possibility of flexible management of computational resources in a *IaaS-like* manner.
Our model can be classified as a Platform-as-a-Service for extending Grid environments with elastic (i.e., virtual) resources. By using **GaaS**, *"privileged"* grid users, i.e. the VO administrator, can define new Grid sites, add computational resources to existing Grid sites and modify the resources aggregation scheme, i.e., site queues.

GaaS services

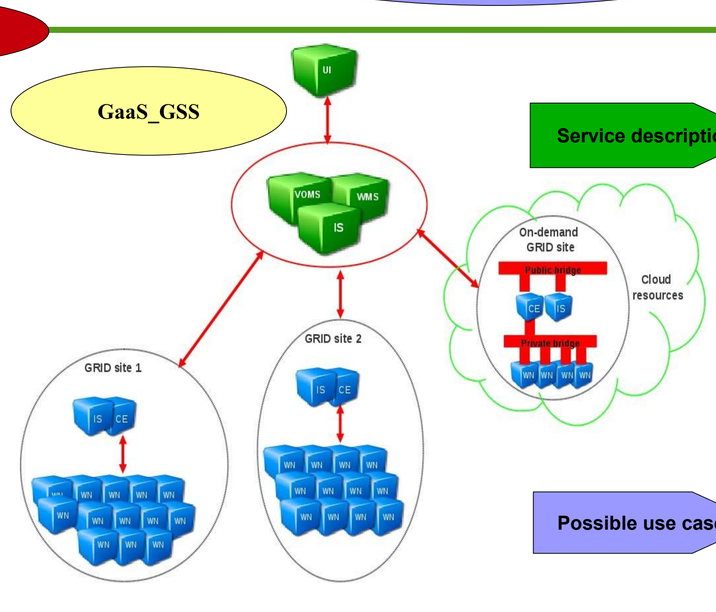
GaaS_WNS



Service description: definition, addition and deletion of WN to be used by the Grid infrastructure;

Possible use cases: realization of efficient datacenters management in term of energy consumption reduction

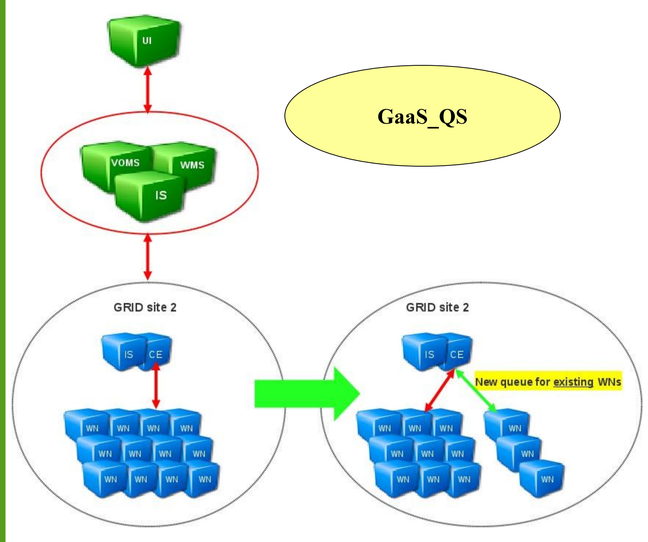
GaaS_GSS



Service description: creation and management of new Grid Sites for an existing VO

Possible use cases: sharing of resources with a community for the life time of a project, to avoid the burden of configuring from scratch all the required services and resources

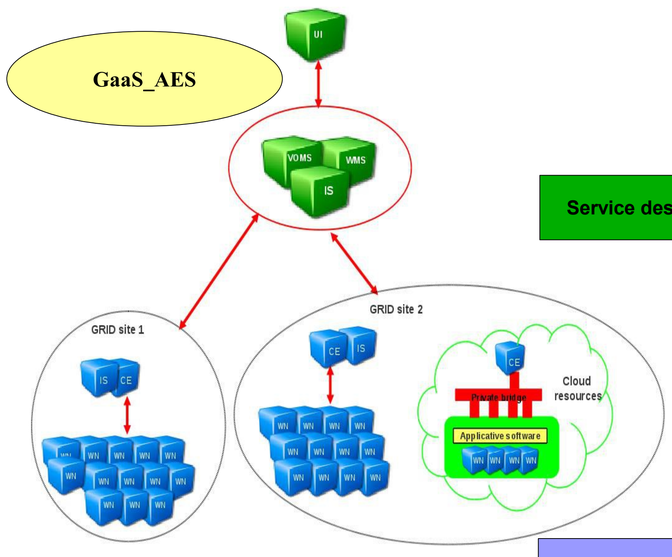
GaaS_QS



Service description: aggregation of existing computing resources in a new queue with different policies

Possible use cases: deployment of particular queue policies (i.e. queue priority for the community users, different values for max walltime of execution, etc.)

GaaS_AES



Service description: creation of a suited runtime environment for a set of applications on existing or new computing resources

Possible use cases: giving the chance to a community to run an application in a proper software enviroment (proper applicative middleware)

The case study for the protoype deployment

The implemented prototype is integrated into the context of the S.Co.P.E. datacenter at the University of Naples Federico II, a self contained grid infrastructure that offers storage and computational resources and all the high level core services for infrastructure management (VOMS, WMS, IS, etc.).

S.Co.P.E. resources are integrated also into national IGI and international EGI relevant distributed computational infrastructures and used from people belonging to different scientific research fields and to VOs from very relevant international experiments (i.e. LHC, ATLAS, Super-B, etc.)

The GaaS prototype, deployed on the S.Co.P.E. production infrastructure at the University of Naples, is based on the **gLite-EMI** middleware and on the **OpenNebula** cloud management system with **Xen hypervisor** to create virtual machines (VM) that host WNs, CE and IS.

GaaS prototype characteristics

To optimize the infrastructural resources used during the provisioning process, and to reduce the overall provisioning time, we took into account the peculiarities of the GaaS system. In particular, we made the following observations:

1) all the VMs are prescribed to host the same operating system, which is imposed by the gLite-EMI middleware;
2) all the VMs hosting the Grid services (i.e., WN, CE, etc.) can be produced by customizing the configuration of a single VM template.

Computational resources can be both virtual and physical. The use of virtual resources is not denied to HPC users but, if virtual resources are chosen, users are notified by IS, about the possible performance limitations.

Conclusions and future works

We presented **GaaS**, a PaaS model for Grid Computing systems, that lets VO administrators to dynamically customize the grid environment they are offering to VO's unprivileged members. VO administrators can define new Grid Sites, add computational resources to Grid Sites and modify the resources aggregation scheme (queues).

We implemented a prototype of our model and deployed it in a real-world Grid Datacenter. Moreover, our prototype implements a virtual resources fast provisioning scheme, that exploits some properties of the Grid environment.

The presented work is a successful proof-of-concept but many issues still have to be solved.
In particular:
1) we have to assess the applicability of virtualized resources in HPC contexts, the payed overhead, and the possibility to extend the model to a mix of virtualized and physical resources according to the users needs.
2) we are working on solutions able to allow new communities wishing to use the grid to instantiate new grid infrastructure also for the non existing VOs.
3) we are also planning an evaluation of the impact on management operations and costs of our approach, in order to integrate a smart management

Acknoledgments

This work is part of the activities of a multidisciplinary group (GTT) that is responsible for the S.Co.P.E. infrastructure management. The work is also part of the activities carried out by GTT in the context of the Italian Grid Infrastructure (IGI).

References

- I. Foster, *Whats faster-a supercomputer or EC2?*, 2009.
- S.K. Garg and C.S. Yeo and A. Anandasivam and R. Buyya, *Energy-Efficient Scheduling of HPC Applications in Cloud Computing Environments,* 2009.
- D. Nurmi and R. Wolski and C. Grzegorczyk and G. Obertelli and S. Soman and L. Youseff and D. Zagorodnov, *The Eucalyptus Open-source Cloud-computing System*, in Proceedings of 9th IEEE International Symposium on Cluster Computing and the Grid, 2009.
- R. Talaber and T. Brey and L. Lamers, *Using Virtualization to Improve Data Center Efficiency*, The Green Grid White Paper n.19, 2009.
- Barham, P. and Dragovic, B. and Fraser, K. and Hand, S. and Harris, T. and Ho, A. and Neugebauer, R. and Pratt, I. and Warfield, A., *Xen and the art of virtualization*, ACM SIGOPS Operating Systems Review, volume 37, number 5, 164-177, 2003.
- L. Merola on behalf of the SCoPE project, *"The S.Co.P.E. Project"*, Final Workshop of Grid Projects "PON RICERCA 2000-2006, AVVISO 1575", February 2009.