The ENEA gateway approach to provide EGEE/gLite access to unsupported platforms and operating systems

G. Bracco , S. Migliori, A. Quintiliani, A. Santoro, C. Sciò *

ENEA-FIM, ENEA C.R. Frascati, 00044 Frascati (Roma) Italy, (*) Esse3Esse

Summary

The success of the GRID depends also on its flexibility in accommodating the computational resources available over the network. A big effort is under way to develop accepted GRID standards but in the meanwhile solutions have to be found to include into EGEE infrastructure resources based on platforms or operation systems which are not currently supported by gLite middleware.

The ENEA gateway approach provides a working solution to this issue, now enabling GRID access to its AIX SP systems.

ENEA, the Italian agency for the energy, environment and new technologies, has a substantial experience in GRID technologies and its multi-platform HPC resources are integrated in the ENEA-GRID infrastructure.

ENEA participation in EGEE has focused on the interoperability between EGEE and ENEA-GRID and resulted in the development of a gateway architecture. The gateway provides a flexible and affordable solution for the access in principle to all the platforms and operating systems available in ENEA-GRID and has been finalized to the case of the AIX SP system, but tests have also been performed for Altix IA64, IRIX, MacOS X and Solaris.

This result can be used to expand the EGEE GRID capability by including a wider range of resources but also, on the other hand, to take advantage on the maturity of the gLite grid services to offer a working GRID solution to communities that have been up to now discouraged by the middle-ware rigidity.

The poster describes the architecture and the implementation of the gateway solution built on the main components of ENEA-GRID middle-ware, which is based on very mature and reliable software, namely the AFS distributed file system and LSF Multicluster.

The key element of the architecture is a set of Linux proxy machines, running standard gLite middle-ware, which support the communication between the non standard worker nodes and the EGEE infrastructure.

In the last year the ENEA-INFO EGEE site has been certified in the gateway configuration for AIX resources and it is open to production jobs. The site supports at present several VOs (COMPCHEM, EGRID, FUSION) and the experimentation with applications is under way.

Proposals by other interested VO are well accepted!

Two EGEE technical notes have been prepared to document the gateway implementation:

EGEE Technical Note EGEE-TR-2007-001
"The gateway approach providing EGEE/gLite access to non-standard architectures" Bracco, G; Migliori, S; Sciò, C. ; Santoro , A.;
http://doc.cern.ch//archive/electronic/egee/tr/egee-tr-2007-001.pdf

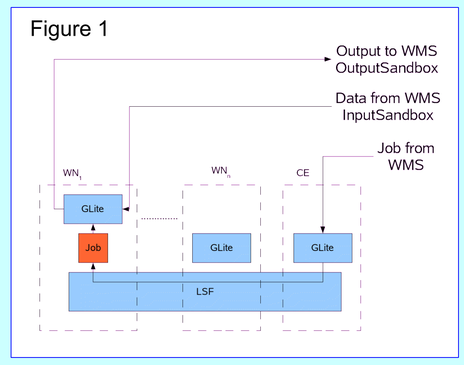
EGEE Technical Note EGEE-TR-2006-006
"AFS Pool Account Users - GSSKLOG and LCMAPS extension to support AFS users as EGEE pool account users"
Bracco, G; Giammarino, L; Migliori, S; Sciò, C.;
http://doc.cern.ch//archive/electronic/egee/tr/egee-tr-2006-006.pdf

The gateway implementation

The Computing Element (CE) used in a standard gLite installation and its relation with the Worker Nodes (WN) and the rest of the EGEE GRID is shown in the Figure 1.
When the Workload Management Service (WMS) sends the job to the CE, the gLite software on the CE employs the resource manager (LSF for ENEA-INFO) to schedule jobs for the various Worker Nodes.
When the job is dispatched to the proper worker node (WN_1), but before it is actually executed, the worker node employs the gLite software installed on itself to setup the job environment (it loads from the WMS storage the files needed to run, known as the InputSandbox). Analogously, after the job execution the Worker Node employs gLite software to store on the WMS storage the output of the computation (the OutputSandbox).
The problem is that this architecture is based on the assumption underlying the EGEE design that all the machines, CE and WN alike, employ the same architecture. In the current version of gLite (3.0.1) the software is written for intel-compatible hardware running Scientific Linux.

CE & WN layout for the standard site

Figure 1



Target for the gateway implementation: no middleware installed on WN

The basic design principle of the ENEA-INFO gateway to EGEE is outlined in Figure 2 and it exploits the presence of AFS shared file system. When the CE receives a job from the WMS, the gLite software on the CE employs LSF to schedule jobs for the various Worker Nodes, as in the standard gLite architecture.

However the worker node is not capable to run the gLite software that recovers the InputSandbox. To solve this problem the LSF configuration has been modified so that **any attempt to execute gLite software on a Worker Node actually executes the command on a specific machine, labeled Proxy Worker Node** which is able to run standard gLite.

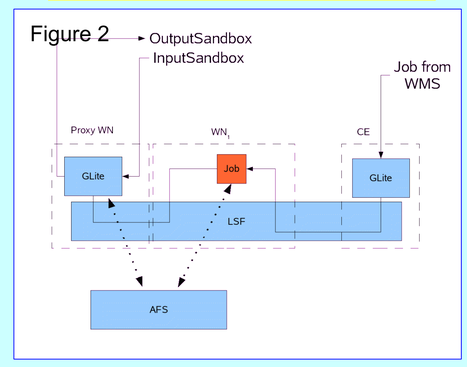
By redirecting the gLite command to the Proxy WN, the command is executed, and the InputSandbox is downloaded into the working directory of the Proxy WN.

The working directory of each grid user is maintained into AFS, and is shared among all the Worker Nodes and the Proxy WN, thus downloading a file into the working directory of the Proxy WN makes it available to all the other Worker Nodes as well. Now the job on the WN_1 can run since its InputSandbox has been correctly downloaded into its working directory. When the job generates output files the OutputSandbox is sent back to the WMS storage by using the same method.

In the above architecture, the Proxy WN may become a bottleneck since its task is to perform requests coming from many Worker Nodes. In that case **a pool of Proxy WN** can be allocated to distribute the load equally among them.

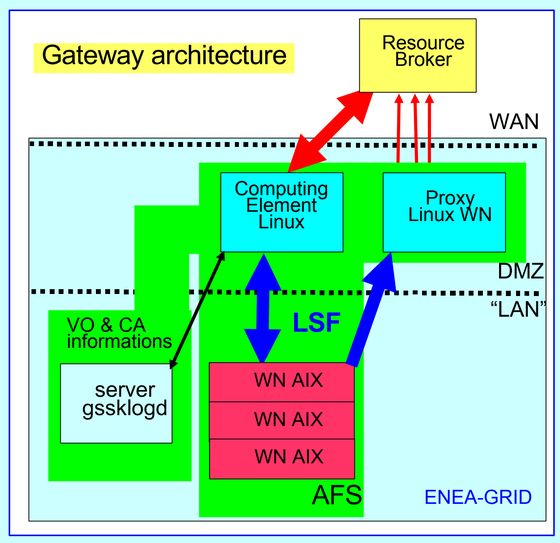
CE & WN layout for the gateway

Figure 2



Authentication & Authorization issues

In a standard EGEE site GRID users are mapped to the local UNIX users by the LCAS/LCMAPS components of gLite middle-ware while ENEA-GRID users are managed using AFS resources and Kerberos 5 authentication so a compatibility solution has to be found.
LCAS/LCMAPS packages provide some integration for AFS users but not in a sufficient way for this implementation, so that a patched version of the packages has been implemented.
Moreover EGEE authentication is based on X509 certificates, which have been extended to incorporate Virtual Organization information using VOMS system. While AFS and X509 compatibility is managed by the standard **gssklog package** a development was required to add support also to VO extension.
AFS file system is also used to share the required informations between the CE and the host of the gssklogd service (k5start has been used to obtain the AFS tokens for the ad-hoc cron jobs).



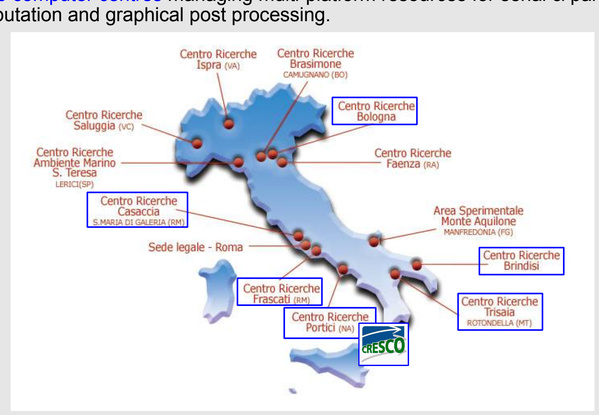
Modified gLite components

YAIM: config_nfs_sw_dir_server, config_nfs_sw_dir_client, config_users
Gatekeeper: lsf.pm, cleanup-grid-accounts.sh
Information system: lcg-info-dynamic-lsf
Worker nodes: the commands have been wrapped for a remote execution on the Proxy Worker Node by means of the **lsrun** command of LSF.

ENEA

[Italian National Agency for New Technologies, Energy and Environment]

12 Research sites and a Central Computer and Network Service (ENEA-INFO) with 6 computer centres managing multi-platform resources for serial & parallel computation and graphical post processing.



ENEA GRID architecture

GRID functionalities (unique authentication, authorization, resource access and resource discovery) are provided using "mature", multi-platform components:

Distributed File System: OpenAFS

Resource Manager: LSF Multicluster [www.platform.com]

Unified user interface: Java & Citrix Technologies

These components constitute the ENEA-GRID Middleware.

OpenAFS

- user homes, software and data distribution
- integration with LSF
- user authentication/authorization, Kerberos V

ENEA GRID

ENEA-GRID computational resources:

- Hardware: ~100 hosts and ~650 cpu : IBM SP; SGI Altix & Onyx; Linux clusters 32/ia64/x86_64; Apple cluster; Windows servers. Most relevant resources:
IBM SP5 258 cpu; 3 frames of IBM SP4 96 cpu
- software: commercial codes (fluent, ansys, abaqus..); elaboration environments (Matlab, IDL..)

ENEA GRID mission [started 1999]:

- provide a unified user environment and an homogeneous access method for all ENEA researchers, irrespective of their location.
- optimize the utilization of the available resources

CRESCO HPC Centre

www.cresco.enea.it

CRESCO (Computational Research Center for Complex Systems) is an ENEA Project, co-funded by the Italian Ministry of University and Research (MUR). The project will be functionally built around a HPC platform and 3 scientific thematic laboratories:

- the Computing Science Laboratory, hosting activities on HW and SW design, GRID technology and HPC platform management

The HPC system (installation 1Q 2008) will consist of a ~2500 cores (x86_64) resource (~25 Tflops peak), InfiniBand connected with a 120 TB storage area. The resource, part of ENEA-GRID, will be made available to EGEE GRID using gLite middle-ware through the gateway approach.

- the Computational Systems Biology Laboratory, with activities in the Life Science domain, ranging from the "post-omic" sciences (genomics, interactomics, metabolomics) to Systems Biology;
- the Complex Networks Systems Laboratory, hosting activities on complex technological infrastructures, for the analysis of Large National Critical Infrastructures.

Issues

The gateway implementation has some limitations, due to the unavailability of the middleware on the Worker Nodes. The Worker Node API are not available and also the monitoring is partially implemented.

As a result, RGMA is not available as also the Worker Node GRIDICE components. A work around solution can be found for GRIDICE, by collecting the required information directly using a dedicated script on the information collecting machine, by means of native LSF commands.

Conclusion

The ENEA-INFO site has been certified for the production grid service providing access both to linux and AIX Worker Nodes.

GOC/GSTAT page with AIX WN information