

Trieste, Italy April 10-11 2008

The ENEA Approach to Interoperability

A. Santoro, C. Scio^{*}, G. Bracco, S. Migliori, A. Quintiliani, A. Rocchi

ENEA FIM, C.R. ENEA Frascati
V. E. Fermi 45 Frascati ROMA (Italy)
bracco@frascati.enea.it
santoro@frascati.enea.it

(*) Esse3Esse, Roma, Italy

Motivations

- Currently standard EGEE infrastructure offers only resources based on HW/OS platforms supported by gLite (IA32/x86_64/SL).
- SPAGO (Shared Proxy Approach for GRID Objects) concept now provides access from EGEE to the ENEA AIX SP systems, belonging to ENEA-GRID.
 - Implementation based on ENEA-GRID middleware, namely OpenAFS and LSF Multicluster.
- This presentation: a simpler solution for any standard UNIX/Posix
 - NFS and ssh can be adopted instead of OpenAFS and LSF Multicluster

ENEA

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

12 Research sites and 6 computer centres with multi-platform resources integrated in ENEA-GRID, a GRID-oriented infrastructure based on LSF Multicluster and OpenAFS WAN file system. (4 Tflops now, mainly AIX, 25 Tflops Linux are coming)



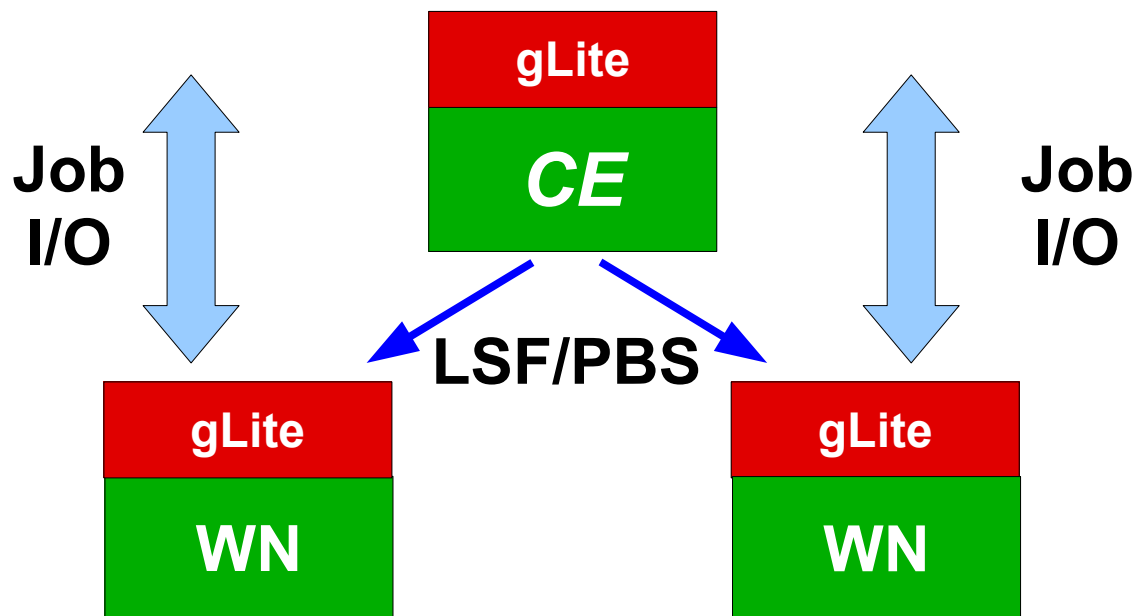
ENEA GRID middleware:

- LSF Multicluster
- OpenAFS File System

ENEA contribution to EGEE

- Research project focused on offering AIX resources to EGEE
 - Interoperability ENEA-GRID/EGEE
- ENEA-INFO EGEE site: three CEs with different resources
 - 1) Cluster of 14 Intel nodes running Linux
 - 2) Our role in EGEE/EGEE2: 105 **G4 CPUs** (4 SP4 frames) running **AIX 5.2**
 - 3) **THIS PRESENTATION**: cluster of 10 Intel nodes with Linux
 - Same interoperability technique in 2) using more common instruments (NFS/SSH)

gLite Resource Restrictions

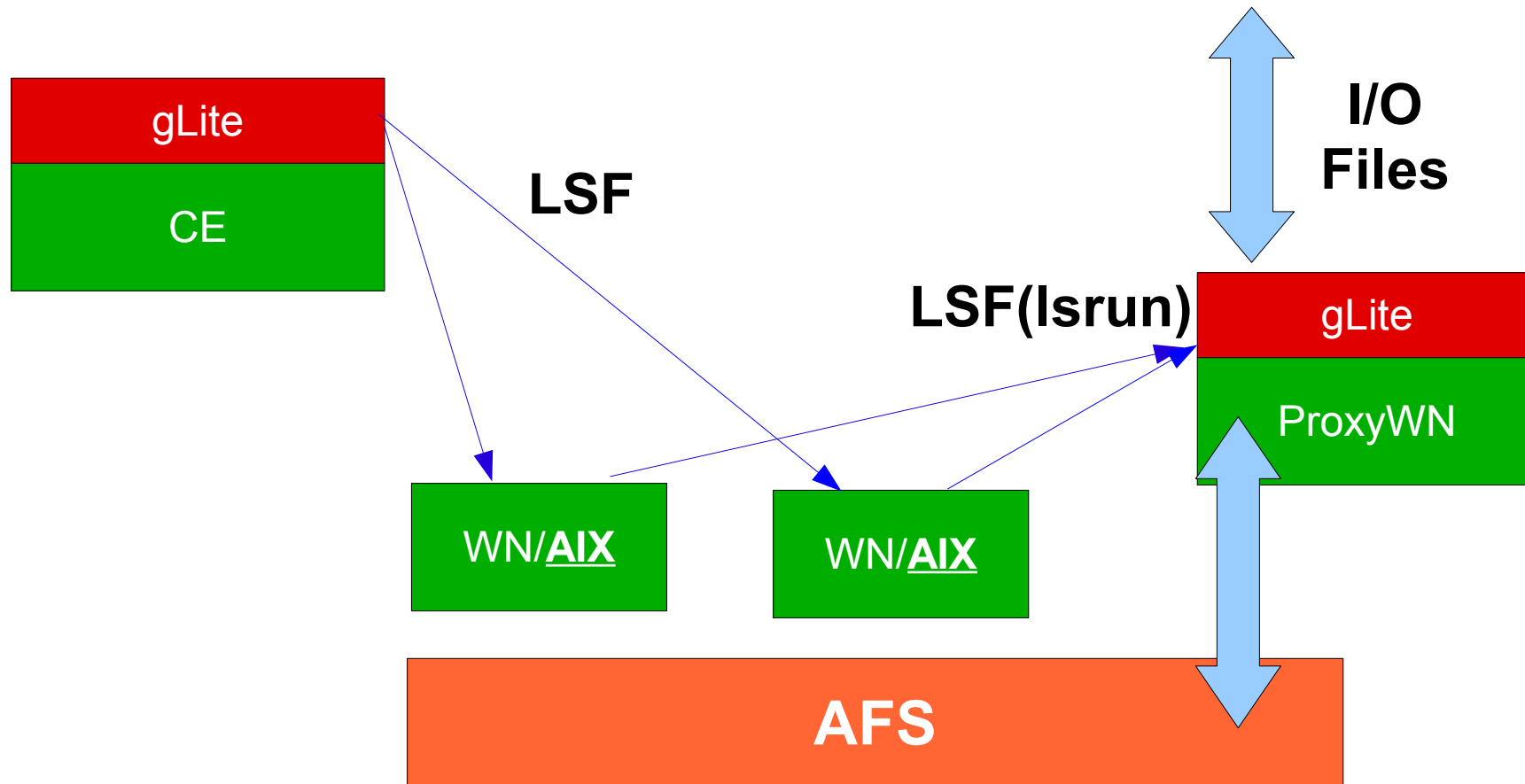


Limitations:

- gLite must be installed on each WN → Only intel/SL machines participate to EGEE
- WNs must directly communicate with the RB → security/firewall management issues

Serious restrictions on computational resources available!

Shared Proxy Approach for GRID Objects (SPAGO)



- Worker nodes do not mount gLite
- Required additional gLite machine (“proxy”)

Redirection to ProxyWN through wrappers

gLite Standard

- **WN: JOBSRIPT**

- Globus-url-copy input-file
- JOB
- Globus-url-copy output-file

SPAGO

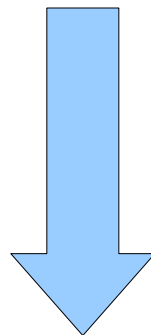
- **WN: JOBSRIPT**

- Globus-url-copy input-file
 - Isrun -m proxy Globus-url-copy input-file
- JOB
- Globus-url-copy output-file
 - Isrun -m proxy Globus-url-copy output-file

→ **Isrun**: prompt job execution

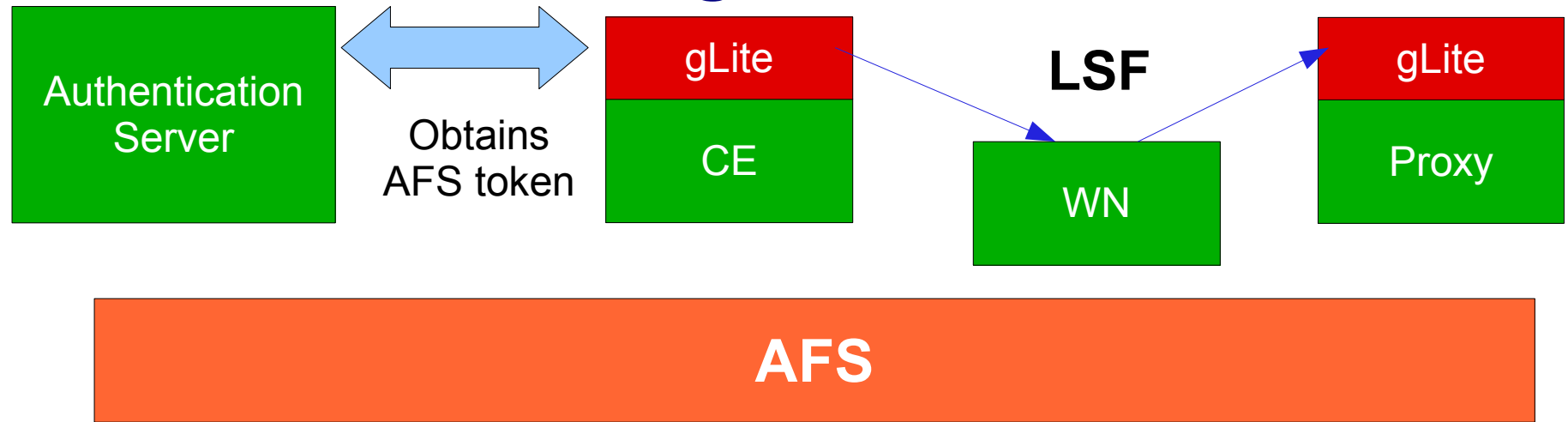
Authentication Issues

- EGEE employs X509 certificates
- ENEA-GRID employs Kerberos tickets and AFS tokens
- Conversion from X509 to kerberos users is built-in into gLite.



Access to AFS is not given

Obtaining AFS tokens



1. CE periodically updates the authentication information on AFS
2. CE receives a job from EGEE user A
3. CE obtains a token for user A from authentication server
4. CE sends job to WN using LSF, propagating the token
5. WN may use the proxy through LSF, propagating the token

Porting SPAGO outside ENEA-GRID?

Requirements in ENEA-GRID:

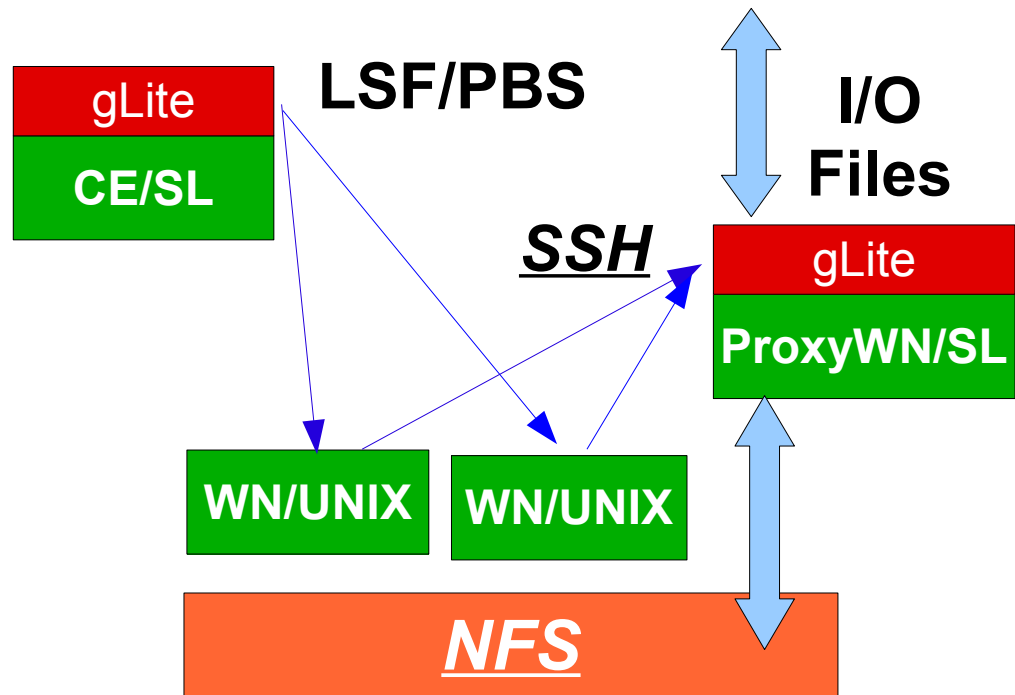
- **AFS** (WAN Shared Filesystem)
 - Authentication interoperability problem (X509 vs AFS token):
Solved
- **LSF** (Job dispatching mechanism)

Analogous tools available in common UNIX systems

- Shared filesystem: **NFS**
- Elementary job dispatching mechanism: **SSH** (*in place of **lrun***)

This simplified version of SPAGO allows grid sites to share any kind of UNIX-based computational resource

SPAGO Approach for a typical UNIX system



Real Case Implementation

- ENEA Fusion department
 - 10 CPUs with SPAGO
 - Employ SSH/NFS
 - Support VO “fusion”

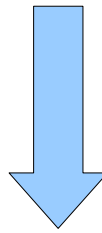
Any typical Shared Filesystem (e.g. GPFS, LUSTRE) and remote execution mechanism (e.g. RSH) can be employed

What to do for a NFS/SSH SPAGO approach

- **Account management separated from Glite**
 - Grid and pool accounts home directories must be located on NFS
- **Wrapper generation**
 - We use a tool for SSH. It is possible to write plugins for other dispatchers
- **Cleanup-grid-accounts turned on only on ProxyWN or CE**
- **SSH configuration**
 - Server and Client configuration for host-based authentication
- **VO public certificates copied on shared file system**
- **Environment variables on WNs**
 - X509_CERT_DIR
 - LOCATION environment variables (e.g. GLOBUS_LOCATION)
- **(optional) Dispatcher using SSH must be prepared.**
 - Required only in case of more than one proxy (to support scalability)

Account Management

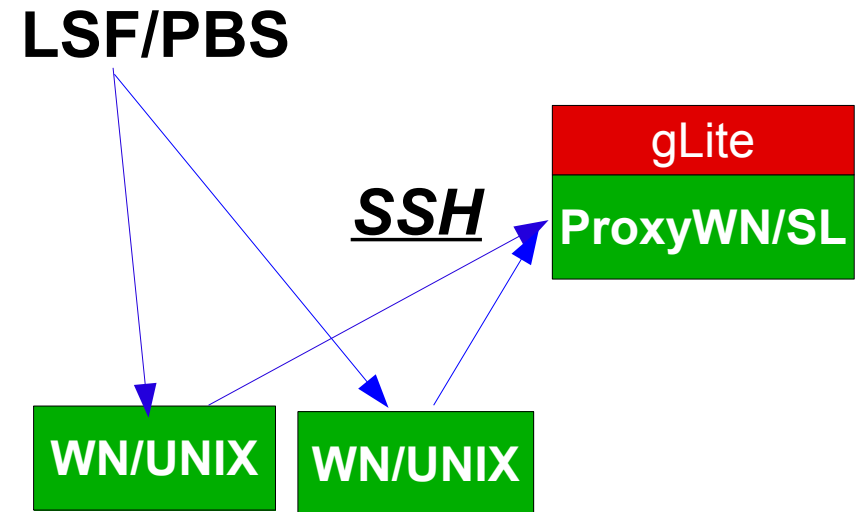
- Grid and pool accounts home directories must be located on NFS
- Cleanup-grid-accounts turned on only on ProxyWN or CE



- Account Management external (i.e. NIS)
 - We use an ad-hoc script to generate and disseminate common accounts
- “config_users” script of gLite must be modified
 - Accounts are already created, they do not need to be recreated
 - Cleanup-grid-accounts must be turned on only on CE

Wrapper Generation

- Wrappers are used to redirect requests from LSF/PBS to the ProxyWN



- We use a tool supporting SSH and LFS. It generates:
 - Redirection wrappers for any WN command in a list.
 - Wrappers “Not supported” for commands that MUST run on WN.
- It is possible to write other plugins for other dispatchers

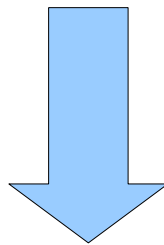
SSH Configuration: host-based authentication

- **Client Configuration**
 - ssh_config on each WN
 - HostbasedAuthentication yes
 - EnableSSHKeysign yes
 - ssh-keysign must be set as 4755 (SUID privileges)
 - Server must be in ssh_known_hosts

- **Server Configuration on ProxyWN**
 - sshd_config
 - HostAuthentication yes
 - shosts.equiv must contain all the Wns.

Public Certificates on Shared Filesystem

- WN might need to check the validity of certificates coming with the job.
- VO and CA public certificates must be readable by WNs.
- Typically they reside on local directories of each grid WN (part of gLite distribution).



- They must be copied on the shared filesystem to be readable by WNs.
- CRON script keeps CE certificates aligned with those in shared filesystem

Setting Environment Variables on WN

- **X509_CERT_DIR**
Should point to directory holding certificates on shared filesystem
- **“_LOCATION” and “_PATH” (e.g. GLOBUS_LOCATION)**
Should point to directories holding generated wrappers
- **Loaded in environment using:**
 - LSF: lsf_jobstarter
 - SSH: .ssh/environment file
 - Exception: “GLOBUS_LOCATION” of CE overrides any setting.
 - Must be overridden by modifying the lsf.pm script
- **Currently hand-managed**
- **Working on using environment of WN tarballs.**

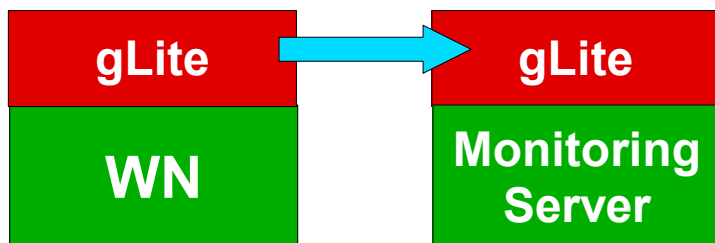
Dispatcher Using SSH (*optional*)

- Scalability issues suggest to use more than one ProxyWN.
- Dispatching mechanism needed to choose the right ProxyWN.
- Required only in case of more than one proxy (to support scalability)
- Possibilities:
 - LSF
 - Bash script implementing RR dispatching.

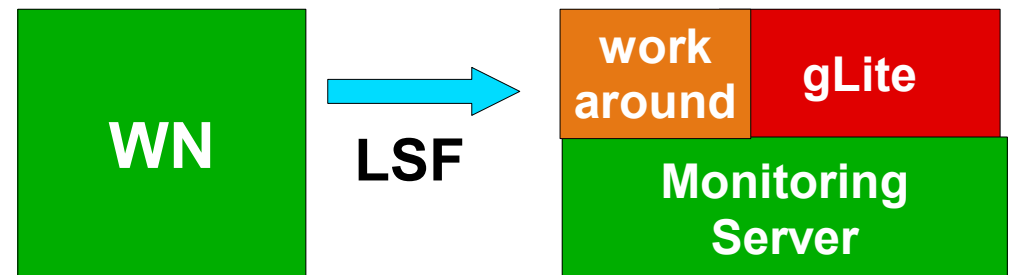
SPAGO Limitation

- Sometimes EGEE requires additional services on WN
 - RGMA (....)
 - GridICE (monitoring system)
- It is often possible to find a work-around
 - eg: GridICE - *Already in production in ENEA-INFO EGEE site*

Typical GridICE



Modified GridICE



Suggestions for a SPAGO-friendly development

- **Architecture-dependent modules should not run on WNs.**
 - Some modules (like GridICE and RGMA) have WN-resident modules.
 - Modules running on WN should explicitly state whether:
 - They are architecture dependent.
 - They must be run on WN or they can be moved on a Proxy.
- **Environment variables on CE and WN should be separated**
 - A few main environment variables of CE override environment variables of the WN.
- **A “SPAGO” tag in the Information System to identify SPAGO-based systems**

References for ENEA-GRID implementation

Technical Reports

- 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>

Recent Presentations

- Grid Camp 2007, WLCG Tier 2 & EU-IndiaGrid Workshop, Taipei 28/10/2007-2/11/2007, "The ENEA gateway approach to provide EGEE/gLite access to non-standard architectures" Bracco, G; Migliori S.; Quintiliani A. ; S. Sciò, C. ; Santoro, A.; reported by G. Andronico in the "Interoperability in Asia" session.
- EGEE'07 Conference Budapest (Hungary) 1-5/10/2007, poster & demo session. "The ENEA gateway approach to provide EGEE/gLite access to non-standard architectures" Bracco, G; Migliori S.; Quintiliani A. ; S. Sciò, C. ; Santoro , A.; (abstract), poster id 135, stand #12.

Wrapping up...

- **For gLite Site Admins**
 - No need to offer only WN based on IA32/x86_64/SL platform: any UNIX-like WN can be supported using SPAGO.
 - e.g. we plan to bring into EGEE a CE with 2500 x86_64 cores based on RH 5.1 (CRESCO project, #180 in the top 500 supercomputer list).
- **For gLite users**
 - It is possible to run software not specifically designed for IA32/x86_64/SL.
- **Feel free to contact us:**
 - if you have to support non-standard UNIX system in gLite.
 - if you have AIX code that should be run in EGEE infrastructure.
 - <http://www.afs.enea.it/project/eneaegee/>