

Examples of applications ported on grid



Emidio Giorgio
INFN Catania
EGEE NA3

emidio.giorgio@ct.infn.it

Tutorial CYCLOPS

CNAF April 13th, 2007



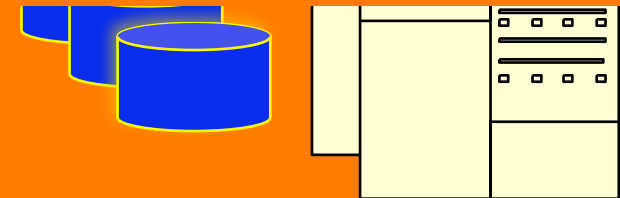
Grid and Application Development

- Application development in Grid implies exploitation of APIs, tools and environment providing Grid capabilities in order to perform complex tasks and achieve diverse goals.

Application

application toolkits

higher level grid services (brokering...)
Basic grid services : AA, job submission, info



Computational Grids



GridApplications



Job concept

- **gLite/LCG middleware follows the job submission concept for application execution and resource sharing.**
- **A job is a self contained entity that packages and conveys all the required information and artifacts for the successful remote execution of an application.**
 - Executable files
 - Input/Output data
 - Parameters
 - Environment
 - Infrastructure Requirements
 - Workflows





Type of applications....

- The work done to port the application depends on its characteristics, complexity and the level of interactions with external grid services
- Size of input/output data
- Metadata request
- Access to external software (database, libraries...)





Type of “Gridification”

- **No development.** Wrap existing applications as jobs. No source code modification is required
- **Minor modifications.** The application exposes minimal interaction with the grid services (e.g. Data Managements)
- **Major modifications.** A wide portion of the code is rewritten to adopt to the new environment (e.g. parallelization, metadata, information)
- **Pure grid applications.** Developed from scratch. Extensively exploit existing grid services to provide new capabilities customized for a specific domain (e.g. metadata, job management, credential management)





No development

- It is the simplest case : the application has not external dependencies, just needs computational power
- Being sure that the application runs on the Worker Node, the executable, and the input files it may need are carried in the Input Sandbox and therefore executed
- Although the application can be directly run, it's recommended to invoke it from a supplied script which prepare the environment before the execution

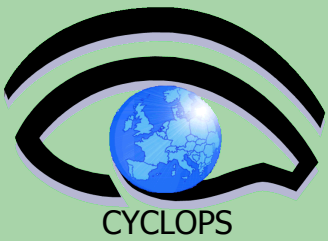




Minor modifications

- The application has minimal interactions with grid service : it is self consistent but...
 - e.g. takes input files from an SE
 - output files are stored on a SE
 - Perform I/O from a metadata catalog
 - the application itself can be downloaded from a SE
- User provides an adapter script which executes needed tasks, set the environment and run the application





Major modifications

- A part of the application software is rewritten taking care of the new environment : e.g. because of parallelization, interaction at run time with grid services...
 - Use of API
 - Web Service invoked
 - CLI wrap
- Source code recompiled taking into account this





Pure grid application

- Entirely developed from scratch taking into account grid capabilities
- Better exploits grid facilities
- Needs a deep knowledge of grid environment and possibilities in order to engineer entirely the application logic over a grid environment
- Very rare to meet...



Additional software installed

- Independently from the integration level, applications may need external software (libraries) that, as application itself, can be provided in several ways
 - at run time
 - carrying out it within IS
 - Downloading it from a Storage Element
 - permanent install on the CE
 - as tar.gz installed by the VO software manager
 - as rpm being part of the release



Software installation on CE

- The Experiment Software Manager (ESM) is the member of the experiment VO entitled to install Application Software in the different sites.
- The ESM is generally acknowledged through a VOMS role (SoftwareManager)
- The ESM can manage (install, validate, remove...) Experiment Software on a site at any time through a normal Grid job, without previous communication to the site administrators.
- Such job has in general no scheduling priorities and will be treated as any other job of the same VO in the same queue.
- **lcg-asis** is a good generator for such installation jobs





Software installation on CE

- The site provides a dedicated space where each supported VO can install or remove software. The amount of available space must be negotiated between the VO and the site.
- An environmental variable holds the path for the location of a such space. Its format is the following:

VO_<name_of_VO>_SW_DIR





Software installation : validation

- Once the software is installed, the ESM can perform the validation.
- The validation is meant as a process or a series of processes and procedures that verify(ies) the installation of the software.
- It can be performed in the same job, after installation (validation on the fly) or later on, via a dedicated job.





Software validation

- If the ESM judges that the software is correctly installed he publishes in the Information System (IS) the tag which identifies unequivocally the software.
- Such tag is added to the **GlueHostApplicationSoftwareRunTimeEnvironment** attribute of the IS, using the GRIS running in the CE.
- Jobs requesting for a particular piece of software can be directed to the appropriate CE just by setting special requirements on the JDL.





Invoke Application

- **From the UI**
 - Command Line Interfaces / Scripts
 - APIs
 - Higher level tools
- **From portals (like GENIUS developed by NICE srr)**
 - For recurring tasks: “core grid services” as well as application layer
 - Accessible from any browser
 - Tailored to applications





- Thanks to Dr. Valeria Ardizzone for the most of this slides !

