eGee

Enabling Grids for E-sciencE

Practical using WMProxy

advanced job submission

Information Society

- **Please bookmark the agenda page.**
- **You will need to refer to it during the practical.**

- **Browse to:**
- **http://agenda.cern.ch/fullAgenda.php?ida=a061960**

Enabling Grids for E-sciencE

- **We are using the GILDA testbed today**
  - The production EGEE grid looks like this!

- **The practical exercises are to illustrate "how"**
  - Not using typical jobs for running on a grid!!
  - But to show how EGEE grid services are used, jobs are submitted, output retrieved,…

- **We will use the Command-Line Interfaces on a "User Interface" (UI) machine**
  - "UI" is your interface to the GILDA Grid
    - Where your digital credentials are held
    - Client tools are already installed

Enabling Grids for E-sciencE

- **WMProxy (Workload Manager Proxy)**
  - is a new service providing access to the gLite Workload Management System (WMS) functionality through a simple Web Services based interface.
  - has been designed to efficiently handle a large number of requests for job submission and control to the WMS
  - the service interface addresses the Web Services and SOA (Service Oriented Architecture) architecture standards

Enabling Grids for E-sciencE

- **Direct Acyclic Graphs of jobs (DAG): set of jobs where the input, output, or execution of one or more jobs depends on one or more other jobs**

- **Parametric Jobs: they have one or more parametric attributes in the JDL, whose values vary according to a parameter**

- **Job *Collection :* a set of independent jobs that for some reason (known to the user) have to be submitted, monitored and controlled as a single request**

- **JDL has been extended to allow specification of the input sandbox at the level of the compound request (i.e. DAGs, Collections and Parametric jobs)**

- **Input Sandbox can**
  - Be shared among nodes of collection/DAG $\rightarrow$ saving bandwidth use
  - Contain URI pointing to files on a remote gridFTP server

Enabling Grids for E-sciencE

- **Support of parallel jobs is essential for CPU-intensive applications.**

- **Most used library for parallel jobs support is (Message Passing Interface) MPI**

- **At the state of the art, parallel jobs can run inside single Computing Elements (CE) <u>only</u>;**

  - several projects are involved into studies concerning the possibility of executing parallel jobs on Worker Nodes (WNs) belonging to differents CEs.

- **You will:**
  - Submit jobs where Input Sandbox contains files from a Storage Element
  - Run job collections
  - Run parametric jobs
  - Run a DAG job (Workflow)
  - Run a simple mpi job

- **Please limit load on resources by reducing number of jobs submitted.**
- **Please work in pairs**

WMS : References

- **WMS User's Guide**
  - https://edms.cern.ch/file/572489/1/EGEE-JRA1-TEC-572489-WMS-guide-v0-2.pdf

- **WM Proxy quick start**
  - http://trinity.datamat.it/projects/EGEE/wiki/wiki.php?n=WMProxyClient.QuickStart

- **JDL Attributes Specification for WM Proxy**
  - https://edms.cern.ch/file/590869/1/EGEE-JRA1-TEC-590869-JDL-Attributes-v0-8.pdf