Enabling Grids for E-science in Europe

NA4 open meeting
Catania, July 15, 2004

Biomed input to JRA1

Johan Montagnat

Recall from this morning overall session

Enabling Grids for E-science in Europe

- LCG2 infrastructure
  - First “Hello biomed world” job executed on June 23
  - VO hosting and User Interface at CC-IN2P3
  - VO acceptance at CNAF
  - Lot of difficulties related to adding a new VO:
    - Adding new VOs in LCG2 (tedious manual configuration, LCFG crashing new VOs every night...)
    - VO acceptance (political level)
    - VO acceptance (technical level: RB, resources...)
  - Jobs submitted from a biomed UI at IBCP
  - RC to be hosted at CC-IN2P3
- glite prototype
  - First glite job executed on June 17
  - Few testing done today, focus on LCG2
- Biomed requirements available on-line
  - http://egee-na4.ct.infn.it/requirements/

LCG2 infrastructure testing

Enabling Grids for E-science in Europe

- Tested
  - VO management
  - Basic testing: jobs submission, selective job submission, MDS
  - No biomed RC yet – no data management testing
  - Interactive jobs
  - Parallel jobs
- Summary of issues
  - VO registration in RB + LCFG destroying new VO + HOSTNAME environment variable problem
  - Interactive jobs are using stdin/out for communication (condor bypass)
    - not sufficient (outbound connectivity expected)
    - far too slow (bytes per second)
  - Parallel jobs: aborted, cannot plan (no compatible resource)
  - Clock synchronization problems (UI and RB living at different times)
- Report from EDG
  - Lack of data security, batch-oriented system (poorly scale to a large number of short jobs, long waiting time even for short jobs), APIs complexity

LCG2 deployment

Enabling Grids for E-science in Europe

- Deployment experiences
  - Biomed UI at IBCP connected to CNAF RB
  - CE, SE, WNs
  - RB, MDS, BDII
- Main problem encountered
  - Documentation
    - Heterogeneous/dispersed
    - Does not cover all problems encountered
  - Hardware compatibility
    - RH7.3 is obsolete and not supported anymore
    - SCSI disks drivers lacking
    - Network card drivers lacking
    - Failed to perform a PXE installation: fallback on floppies
  - Configuration
    - Too many configuration files
    - Too many repetition in configuration
  - Difficulty (impossibility!) of installing several services on a single host

glite testing

Enabling Grids for E-science in Europe

- Tested
  - usage: CERN AFS account (lxplus UI), CERN installation, EGEE VO
  - Standard commands
    - login, register, add, cat, cd, cp, echo, exit, find, quit, kill, submit, whereis, jobinfo, ps, top, cat, help, get, pwd,rmdir, mkdir, touch
  - Downloading and uploading of data and executables (registration and linking)
  - JDL files and execution of sample applications (rendering, CDSS)
    - executable, StdIn, StdOut, InputFile, InputData, OutputFile, split, arguments
  - scripts for data transfer, split and recovery of results
- Comments
  - coherent view of the system
  - prototype weaknesses

glite testing

Enabling Grids for E-science in Europe

- Problems
  - Instabilities: jobs stacked on queue, abnormal terminations, file registration failures...
  - Long and unpredictable delay on processing
  - Irregular registration results
  - Lack of editing facilities in the glite command (turn around: download + edit + upload)
  - Lack on error feedback
    - add do not check URL validity and produces strange errors
    - very little information on errors related to the JDL syntax
    - many internal error message (not for users)
  - Lack of documentation (especially JDL)
  - Some commands syntax too strict (register requires host name for instance)
  - Should not let all jobs outputs world accessible by default, yet expect to be able to specify which output file should be shared

Security related issues

Enabling Grids for E-science in Europe

- Access control
  - fine grain access control: users + groups
  - service certificates for portals (anonymous login)
  - access right delegation
- Data protection
  - storage encryption + network encryption
  - robustness against storage administrators
- Data anonimization
  - application plugins in storage ressources for data transformation
- Replication target control
- Privacy enforcement
  - data protection
  - hiding computing
  - hiding users
- The level of security achieved will constrain some applications

Summary

Enabling Grids for E-science in Europe

- LCG2 infrastructure
  - VO extensibility
  - parallel jobs
  - batch oriented system: interactive jobs, short jobs
  - lack of data security
  - service certificate (portal)
  - no load testing performed
- LCG2 deployment issues
  - UI/APIs weight
  - RH7.3 obsolete
  - configuration is a nightmare
- glite
  - encouraging but still a prototype (stability, stability...)
  - lack of error feedback/documentation
  - always more: editing in glite console