



Enabling Grids for E-science

EGEE Review: Application Assessment of gLite

M. Lamanna EGEE NA4-HEP and LCG ARDA

www.eu-egee.org



- **EGEE JRA1 (gLite):**
 - Re-engineering of the existing middleware
 - Stable and dependable services to be deployable on a **very fast** growing infrastructure
 - *The scale of the EGEE-LCG infrastructure is already larger than expected*
- **EGEE NA4**
 - Dependable system for the applications needed
 - Cover more use cases than large-scale productions
 - *Large number of concurrent users (HEP analysis)*
 - *Groups of users working in variable topologies (lightweight Biomed VOs)*
 - *New applications → new use cases and possible new requirements*
 - General need of “interactivity”:
 - *Interactive analysis on the Grid*
 - *The Grid as the environment where scientist evolve their applications out of a bag of services and tools (not a “simple” batch system)*

- **EGEE NA4 has 3 lines**
 - HEP, Biomed and “Generic” (“New”) applications
- **How does NA4 contribute to the middleware evolution?**
 - Different complementary approaches:
 - HEP: LCG ARDA project
 - *~100% of the whole NA4-HEP effort*
 - Biomed: parallel approach on the old and the new infrastructures
 - “New” applications: interest to follow the evolution to present a complete scenario to new application
 - The NA4 testing effort
 - Dedicated working group
 - In the first phase: contribution to the general gLite testing effort
 - On longer term: requirements checking

- **LCG contribution**

- Personnel effort:

- EGEE NA4-HEP contribution largely overmatched (4F and 6+ UF)

- Management:

- ARDA coherently managed (same lead) with NA4-HEP
- Same persons sitting in LCG PEB and EGEE PEB

- Key area in LCG

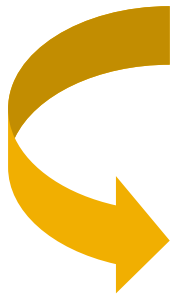
- LCG put a very relevant effort in contributing real experience, extracting general lessons and architectural guidelines:

- *LCG GAG (Grid Advisory Group)*
- *ARDA RTAG (LCG working group)*



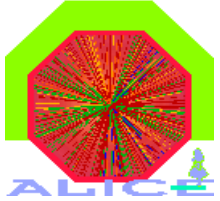
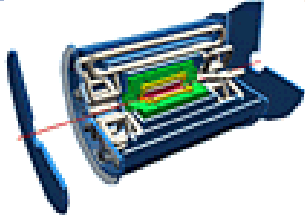

and in the current environment as well (details in the next slide):

- *EGEE PTF*
- *Conferences, workshops...*
- *Contacts with other initiatives (OSG, UK metadata group)*

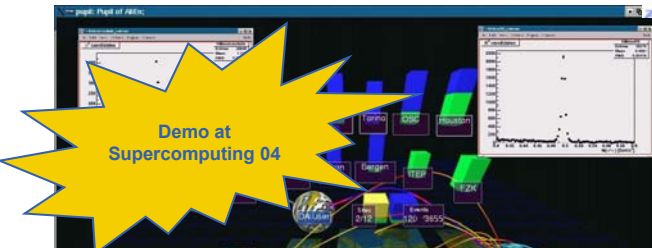
- **ARDA continues on this line:**
 - Started as project in LCG at the same time of EGEE (4/2004)
 - Put in contact the experiment Grid experts with the gLite teams and other EGEE bodies, e.g.:
 - Review of the architecture and design gLite docs
 - *Formal review process*
 - *Invitation of the ARDA experiments contact into PTF discussion*
 - *Informal channels (mailing lists, etc...)*
 - International events and conferences
 - *CHEP04 (Computing for HEP 2004): 2 presentations*
 - *SuperComputing 2004 (October): 2 demos*
 - *IEEE2004 (October): 1 presentation*
 - *Rio: 1 demo*
 - ARDA workshops
 - *June and October workshops*
 - JRA1-3/NA4 workshops and EGEE joint sessions
 - *July workshop (Catania)*
 - *November joint session in the 2nd EGEE conference (Den Haag)*



- **The complexity of the field requires a great care in the phase of middleware evolution and delivery:**
 - Complex (evolving) requirements
 - New use cases to be explored (for HEP: large-scale analysis)
 - Different communities in the loop (LHC experiments, middleware experts from the experiments and other communities providing large middleware stacks (CMS GEOD, US OSG, LHCb Dirac, etc...))
 - The complexity of the experiment-specific is comparable (often larger) of the “generic” one
 - The experiment do require seamless access to a set of sites (computing resources) but the real usage (therefore the benefit for the LHC scientific programme) will come by exploiting the possibility to build their computing systems on a flexible and dependable infrastructure
- **How to progress?**
 - Build end-to-end prototype systems for the experiments to allow end users to perform analysis tasks

LHC Experiment	Main focus	Basic prototype component	Experiment analysis application framework	Middleware
	GUI to Grid	GANGA	DaVinci	
	Interactive analysis	PROOF ROOT	AliROOT	
	High level services	DIAL	Athena	
	Exploit native gLite functionality	Aligned with APROM strategy	ORCA	

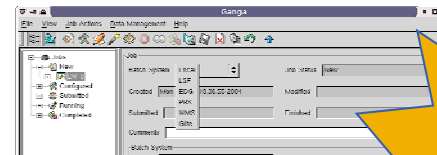
eGEE Interactive Session



Demo at Supercomputing 04

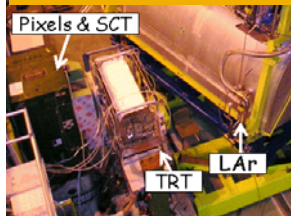
eGEE Current Status

- GANGA job submission handler for gLite is developed
- DaVinci job runs on gLite submitted through GANGA



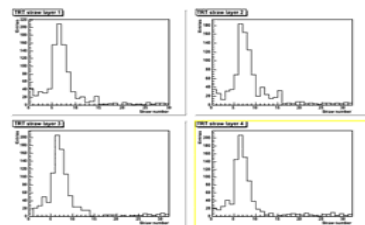
Demo in

All prototypes have been “demoed” within the corresponding users communities

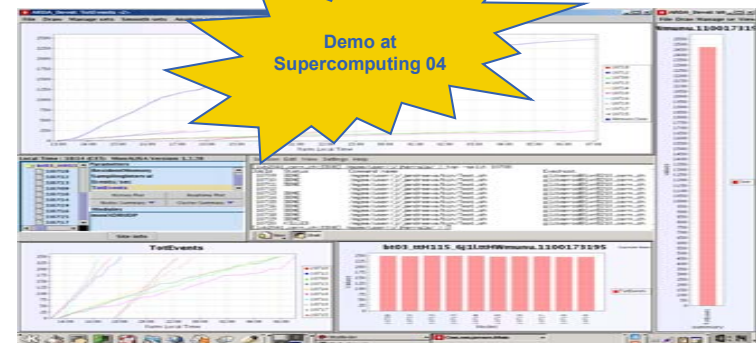


Example:
ATLAS TRT data analysis done by PNPI St Petersburg
Number of straw hits per layer

Standard Athena for testbeam
Data from CASTOR
Processed on gLite worker node



INFSO-RI-508833



Demo at Supercomputing 04

INFSO-RI-508833

Work Load Management



ARDA has been evaluating two WMSs

- WMS derived from Alien – Task Queue
 - available since April
 - pull model
 - integrated with gLite shell, file catalog and package manager
- WMS derived from EDG
 - available since middle of October
 - currently push model (pull model not yet possible but foreseen)
 - not yet integrated with other gLite components (file catalogue, package manager, gLite shell)

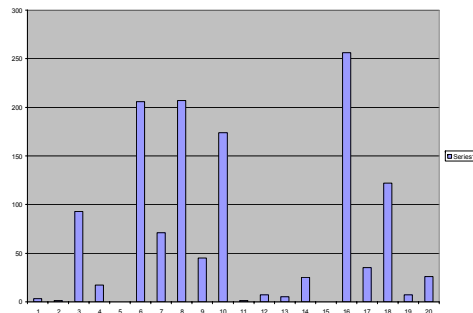
INFSO-RI-508833

gLiteIO tests



- Try to copy 1000 files of 0 to 10KB

DMS: data transfer



- On average an error occurred after 64 Files
- About 10 different error messages observed

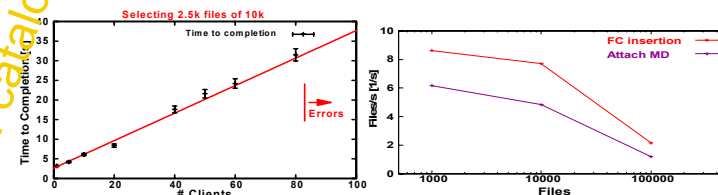
INFSO-RI-508833

Performance



DMS: catalogs


- gLite File catalog
 - Good performance due to streaming
 - 80 concurrent queries, 0.35 s/query, 2.6s startup time



- Fireman catalog
 - First attempt to use the catalog: quite high entrance fee
 - Good performance
 - Not yet stable results due to unexpected crashes
 - We are interacting with the developers

INFSO-RI-508833

Package management



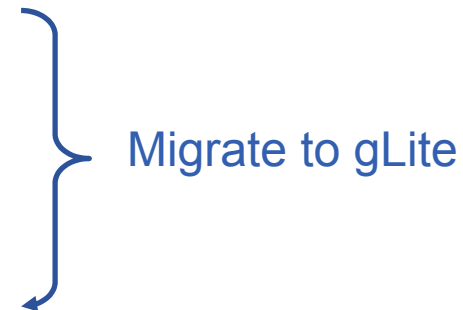
- Multiple approaches exist for handling of the experiment software and user private packages on the Grid
 - Pre-installation of the experiment software is implemented by a site manager with further publishing of the installed software. Job can run only on a site where required package is preinstalled.
 - Installation on demand at the worker node. Installation can be removed as soon as job execution is over.
- Current gLite package management implementation can handle light-weight installations, close to the second approach
- Clearly more work has to be done to satisfy different use cases

INFSO-RI-508833

- **Biomed assessment of gLite**
 - Less effort compared to the NA4-HEP
 - A lot of NA4 activity on the LCG system as well
 - Similar approach
 - Migrate complete applications
 - Specific interests
 - Security
 - *First interesting feature will be available in gLite RC 1.0*
 - *Fine-grained ACL for data access*
 - *Data encryption (at least at some stage of the data handling)*
 - *In Den Haag, very good mutual understanding between NA4 and JRA3 (Security) ☺*
 - Job latency (“short jobs execution” use case)
 - *Quasi-interactive behaviour needed*
 - *As a matter of fact application-level mechanisms could solve it (for Biomed, see gPTM3D; analog approach also found in LHCb DIRAC)*
 - Light weighted VO
 - *Size of collaborations smaller and short lived (compared to HEP)*
 - *Much more dynamics!*

- **Migration of a LCG-2 Application for Volume Rendering on the gLITE Environment.**
 - Volume Rendering of Large Datasets (The Visible Human Project, 512x512x1500x2 bytes).
 - Aimed at Sharing Computing and Memory Resources.
 - Sequential Version Cannot be Executed in a Single Standard Computer due to the High Need of RAM (About 4Gbytes).
 - It Generates Videos with a Rotation Around any Axis.

- **More applications in the pipeline**
 - Running on LCG2: CDDS, gPTM3D
 - Globus Toolkit3: Virtual workbench???
 - AliEn: Mammogrid



- **High interest!**
- **Biomed-specific requirements understood and a lot of activity going on**
 - First key element will be part of gLite RC 1.0
- **Documentation and robustness to be improved**
- **A stable service is a prerequisite for more massive migration effort**

- **Dedicated effort**
 - 2 FTEs (first 8 months) → 3 FTEs (Since PM9)
- **Multiple goals**
 - Requirements verification
 - Contribution to the overall testing effort
 - Coordinated action with JRA1 testing and SA1 certification activities
- **Status**
 - Preparation
 - Mainly done on the LCG system so far
 - Now ready to start to play with gLite as soon the pre-production service become available
- **Specificity**
 - “User-runnable” tests
 - High level test (useful also as tutorial material)

- The “generic application” has the specific feature that they are presenting the grid infrastructure to new applications
 - These applications see the grid now but will become power users when gLite will have delivered its new middleware and advanced functionality
 - Role of portals here is key!

- Genius “isolates” the users (especially new users) from the changing layer

- Interest in shell access

- Absolute necessity of stability!

- During last ARDA Workshop I tried to download, install and configure gLite client **successfully**
- Fedora Core 2 machines
 - the data management related commands worked
 - the job submission commands did not (libstdc++ problem)
- Production testbed UI
 - Data management and job submission worked

But.... Last Saturday data management did not work any more

- **The workshop activity produces a set of open issues we are currently being discussed within NA4 and with external bodies (notably OSG)**
- **Most interesting message:**
 - Natural emersion of a “core application layer”
 - Examples of “common” architectures
 - User
 - Interactive services
 - ...
 - Specific features not forgone
 - Security critical for Biomedicine
- **Re-inject this as input for EGEE PTF**

- **Discovery services:**
 - More or less missing
- **Security related**
 - Encryption in the storage of data (avoid people with local administrator rights to compromise privacy) and in the communication layer
 - ACL granularity
 - Groups / Users
 - Files / Groups of files (“directories”)

- **Compare the experience of DIRAC, gPTM3D, others... with the current plans**
 - Agent-based “mobile systems: in DIRAC these agents are distributed by the LCG2 WMS and then they collect tasks from an experiment-specific service
 - Robustness
 - Handcrafted control channels: in DIRAC, the communication layer is done via instant-message technologies
 - Duplication of effort on the application side?
 - Provisions for interactivity control
 - gPTM3D and ALICE (and maybe others)
 - Response time
 - Quality-of-service and (low) latency requirements

- **gLite provide a usable system (development prototype)**
 - Allow application to preview the system during the evolution and prepare for it
 - Applications provides valuable feedback
 - Test-like experience (bug reports, validation of the fixes)
 - Discussion on the architecture and exposed functionality (now collected in a systematic way in the EGEE PTF)
 - Collect experience, requirements, etc from end user
 - This will continue for advanced-features preview
 - Naturally very small in size/resources (3 sites)
- **NA4 looks forward for the pre-production service**
 - Ultimately delivered by SA1
 - Lot of preparation work/contribution from both JRA1 and NA4
 - Essential to explore
 - Higher complexity: more sites and more end users
 - Larger resources: more CPU/disk (essential to have real users)

- **Applications are the ultimate reason to build infrastructures like LCG and EGEE**
- **gLite assessment is serving multiple purposes and demonstrating the validity of this approach**
 - Increase the readiness of the final application by prototype activity:
 - HEP via ARDA, “Generic” via portals
 - The close loop between application and middleware is a special asset for EGEE
 - Improve the “quality” of the final system
 - All communities active here + testing group
- **Outlook**
 - Preproduction service being prepared (direct input from NA4)
 - gLite development test bed used so far
 - used for full-scale demonstration before the final deployment and feedback on latest features
 - The former activity will be best continued on the preproduction service as soon as available (SA1)
 - The latter will continue on the smaller gLite development prototype (JRA1)