eGee

NA4 contribution to all activity meeting

V. Breton on behalf of NA4

Information Society

Points raised in Q&A note for NA4

- **The user survey must answer the question – what is the benefit the applications see of joining the grid?**

- **Make the requirements the project is working against more public and obvious**

- **Other**
  - need to port/interface (not develop) popular workflow system to gLite to move away from classic batch systems
  - Need to clarify what new applications can expect and need to provide via the MoUs

- **The application users did also not (in all cases) clearly demonstrate the scientific benefits of the EGEE Grid**

Reviewers recommendation to NA4

- **1- Have all current applications migrated to gLite with a very good user satisfaction rating on application development support and grid infrastructure operations**

- **2- Building on the experiences of previous FP5 grid projects, capture full requirements of future user groups, assess needs for new Grid services and plan accordingly for later implementation**

- **3- Clarify the true motivation of users from new application areas right from the beginning. In the extreme, are users from the new areas really interested in applying grid technologies to explore new ways to produce more and better scientific results or simply interested in accessing large and cheap resources?**

Reviewers recommendation for future work

- **Testing and software packaging will be critical to success. Reinforce these also intellectually very demanding activities even further.**

- **Choose new application areas with greatest care: capture user and application requirements in details and determine the impact on Grid infrastructure and services. The decision to deploy any new application (and if yes when) should be based on a very careful analysis of all associated technical and non-technical risks.**

- **The ambition to quasi-fulfill industrial or commercial requirements should be relaxed until the end of the current contract.**

- **If examples can be found of scientific work that could not have been done without the EGEE development (e.g. necessary distributed process, truly enormous calculations), they will be a strong motivation for future development.**

- **General NA4 strategy: difficult to migrate before users are convinced**
  - Show stoppers: robustness and scale (number of sites/nodes)
  - Need to change OS system (RH7.3-> SL)
- **Scale: EGEE-0 and EGEE-1 need to share resources**
  - gLite and LCG-2 should coexist on the same physical machines (worker nodes)
- **General roadmap**
  - Common testing activity will allow to set-up preproduction service
  - Once preproduction service is available, validation of a subset of applications on preproduction service.
  - Once validated, the other applications will migrate

HEP perspective

- **LHC experiments deployed on LCG2**
  - These activities (production, analysis) cannot be stopped since they are critical for the experiments
  - Migration will happen when the new system will be better than the existing one
- **Is gLite better than LCG2?**
  - Common effort with the goal of having gLite on the pre-production service
    - NA4 involved with JRA1 and SA1
  - ARDA contributes with:
    - Common testing effort
    - Introductory informal tutorials
      - *Support for new contributors joining the common testing effort*
      - *Introductory work with LCG2 and experiments experts*

HEP perspective on migration

- ARDA prototype migration
  - Started on the development testbed for all systems
    - *Gain time*
    - *Effectively it is a contribution to the testing and certification*
  - Validation will come from running on the pre-production!
- Validation will need the preproduction service
  - Size is the key: number of CPU, access to all relevant data without major data "copying"
  - Attract new users (a few on the development tests already)
  - LCG2 and experiments experts in the loop (already started)
- ARDA will continue its role
  - Prototypes
  - Contacts with the experiments experts

Biomed migration to gLite

- **Biomed experiments are waiting for a system with at least the same functionalities and robustness as the current middleware.**
  - Last reports mention instability (on the prototype)
  - See the end of March release

- **CDSS (Clinical Decision Support System) first pilot to migrate**
  - UPV actively testing gLite prototype
  - Expecting a service-based infrastructure

- **Migration plans for gPTM3D, GPS@ and remote medecine**
  - no date set yet

- **Other applications wait and see**

- **Some testing started at CINES computing center**

Strategy to deploy gLite on GILDA

- **As soon as v1.0 of gLite will be released together with the instructions for installation and configuration, the new middleware will be deployed to Catania site only and tested**

- **As stated in DNA4.1 and DNA4.2, we will exploit the power of the GENIUS grid portal to ease the transition from LCG-2 to gLite**
  - GENIUS generic services will first be adapted to gLite
  - then, the application specific services of the portal will be modified

- **We will start first with the GILDA "demonstrative" applications so that they can be used immediately in gLite induction courses**

- **After the first test, the new middleware will be deployed on all GILDA sites**

- **For all of this to happen successfully, it is <u>mandatory</u> that LCG-2 and gLite worker nodes can co-exist on the <u>very same</u> physical machines and the access to existing data is guaranteed <u>as promised since the beginning</u>. Otherwise, due to the small scale of GILDA, there would be an intolerable and disrupting interruption of service with strong impact also on NA3 activities**

Strategy to migrate Gen. Apps. to gLite

- As soon as the new features of the middleware will be tested, first WMS services and then DM services will be adopted

- Based on the past successfull experience, the start of the migration of the GENIUS services of the "official" EGEE generic applications from LCG-2 to gLite will be done during retreats between the GILDA team and the experts of the corresponding communities

- The schedule of these retreats will be agreed with the new communities depending on the time scale of their readiness and willingness to adopt the services of the new middleware

- However, tests of the new middleware by the "official" EGEE generic applications communities and the transition from LCG-2 to gLite (to be done in GILDA at the beginning) should by no means delay the process of deployment on the LCG-2/EGEE-0 infrastructure

Testing: recent History

- **Testing Crisis:**
  - Very few functional tests for gLite.
  - gLite cannot be released without a reasonable testsuite.

- **Reorganization**
  - Nick Thackray (SA1) now coordinating JRA1, NA4, and SA1 testing activities.
  - Immediate-term goal is to reproduce the functionality of "Gilbert's Testsuite" for gLite. (25 tests in total)
  - Development, Certification, and Pre-production testbeds "available" for developing the tests (and verifying that they and the software work).

- **ARDA gLite Tutorial**
  - Given at CERN. Excellent overview of gLite functionality.

Status of NA4 test team activity

- **Application Use Case Procedure**
  - Defined for a long time.
  - Have a collection of ~20 application use cases.

- **Test Framework**
  - Python/pyUnit based (agreed with JRA1).
  - Produces output in XML format (agreed with JRA1).
  - Proto-module exists in the EGEE CVS repository.

- **Available (or soon to be) Tests**
  - Already cover majority of "Gilbert's" job submission tests.
  - Porting of these fairly trivial.
  - Group has agreed to provide 4 additional tests by the 18th.

- **See https://edms.cern.ch/document/569648/1.**

Testing: future

- **Crisis Management.**
  - Rene, Eric, and Delphine providing requested tests.
  - They're definitely in the "white hat" role.

- **NA4 Test Cases**
  - Return to these after short-term crisis is over.
  - Still need to be done in rather short order.

- **NA4 mandate(ARDA & NA4 test team) is application level testing**

requirements capture, need assessment
for future implementation

- **Status: Process to collect requirements is set-up and operational**
  - Extension to projects or applications connected to EGEE
    - External projects and user communities have entries into the requirements database

- **Needs:**
  - Collect user requirements more broadly taking advantage of previous FP5 projects
    - Resources needed to evaluate outcome of FP5 projects
  - Define requirements process from their collection to the implementation of grid services
    - How are project requirements taken into account for middleware implementation ?

Clarify the true motivation of users from new application areas

- **New application selection process up to now**
  - EGAAP Initial philosophy is to attract people, not protect project resources
  - Two step process to accept new applications
    - First light filter: EGAAP, acting as a scientific committee
    - Second technical filter and formal approval by PEB

- **Need to further strengthen the selection process**
  - Section 9.11, p15: "Capture user and application requirements in details and determine the impact on Grid infrastructure and services. The decision to deploy any new application (and if yes when) should be based on a very careful analysis of all associated technical and non-technical risks."

- **Careful evaluation of new applications prior to EGAAP approval requires SA1 involvment**

How to demonstrate a better usage of the grids
by scientific communities ? (1/2)

- **On-going activity on deploying complex workflows**
  - Integration of TRIANA in GENIUS (INFN-Catania)
  - Interfacing of TAVERNA with LCG2 (CNRS-I3S)
- **Collaboration with TRIANA group from GridLab is pursuing very well**
  - First full writer/reader of LCG-2 DAG JDL files coded in TRIANA
  - Latest version of TRIANA deployed on GILDA at Catania and integrated in GENIUS
  - Tests are undergoing
- **Taverna: worflow management system (from MyGRID UK e-science project)**
  - Development of WS wrapper to LCG2 command line interface

How to demonstrate a better usage of the grids by scientific communities ? (2/2)

- **Examples of scientific work impossible without a grid to be identified**
    - HEP
    - Large scale docking for drug discovery on tropical diseases
    - …
- **Could EGEE resources be dedicated to non HEP data challenges for short periods of time ?**

MoU: Status

- **Reminder: MoU goals**
  - Identification of the main actors in the different organizations (EGEE NA4, NA3, SA1, Application…) in order to work in collaboration
  - Expression of application needs (CPU, storage, training…) and evaluation of their impacts on EGEE
  - EGEE and application commitments with deadlines on specific tasks

- **Document commented by F.Harris, R.Barbera, G.Wormser, A.Mills, D.Fergusson.**

- **Document sent to PEB on 25/02/05, awaiting for additional comments and** validation

MoU: Roadmap

- **Objective : agreed MoUs with new applications (EGEODE, Comp Chemistry, Earth Sciences (academic and industry), Drug Discovery, Planck, Magic) before Athens**

- **First MoU foreseen with Drug Discovery application: positive initial feedback**
  - Good feeling about the document
  - The document helped them to define and quantify more precisely their needs
  - They appreciate the vision of the roadmap with the commitments from EGEE and from the application partners
  - The chapter on metrics indicates what they must survey

- **Deadlines:**
  - Validation of MoU by PEB = March 14th
  - MoU sent to the applications <18/03/05
  - MoU sent to EGEE (NA4, NA3, SA1) <26/03/05
  - MoU agreed between EGEE and the application <18/04 /05

New applications

- **Next EGAAP meeting**
  - No EGAAP call
  - Need to get in touch with communities invited to come back at last EGAAP meeting
  - Meeting dedicated to survey of application status
  - Invitation to participate to EAC members

- **Future: next call in summer with evaluation in the fall**
  - Need to change approach, for instance document the real services offered by SA1 ?
    - No MPI, no interactivity, no license management ?

Licensing issues

- **Industry forum: not the place to go into technical details, theoretical work**
  - Y. Guerin to be aware of these issues

- **License management is also required by scientific applications (comp. Chem., durg discovery, CGG,...). So licensing issues should be addressed during EGEE.**

- **Need to work on a specific use case:**
  - Example on http://egee-na4.ct.infn.it/genapps/wiki/index.php/UseCasesForLicensedSoftwareAgreement

- **Project wide issue**
  - NA4 contact points: Marcel Soberman, Roberto Barbera and Guy Wormser

NA4 15 month plan

- **No major change to the text already available on EDMS**
  - Revision under way following reviewers feedback
- **Migration to gLite of a subset of applications**
- **Virtuous cycle**
  - Improve application identification and selection process
  - Make sure the user communities already on LCG2 are happy
  - Set up the tools to survey user satisfaction

NA4 15 month plan: issues

- **Issues from present version identified before review**
  - prepare the transition from LCG-2 to gLite foreseen after the first EC review. It is the biggest challenge in light of the applications already active on the infrastructure
  - get reliable indicators of decentralized activity on the production infrastructure as most of the statistics produced so far are based on estimates done by hand.
  - prepare the transition from NA4 supported application deployment to autonomous usage of the infrastructure for the first scientific communities selected by EGAAP.
  - project has underestimated effort in managing the virtuous cycle and this is putting unforeseen demands on time necessary for liaison involving all the actors e.g. NA4, NA1, NA2, NA3, SA1.

- Move MN4.2 Milestone to PM14
- Reduce number of deliverables by suppressing DNA4.3.3

MNA4.2	M12 -> M14	First external review of Applications Identification and Support with feedback
DNA4.3.2	M15 -> M18(19)	First revision of EGEE Application Migration Progress report
DNA4.3.3	M21	Second revision of EGEE Application Migration Progress report
MNA4.3	M24	Second external review of Applications Identification and Support with feedback
DNA4.4 -> DNA4.3.3	M24->M22	Final Report of Application Identification and Support Activity, including Application Migration status

eGee User Survey: Status

- **Goal: Evaluate user satisfaction for milestone MNA4.2**
    - Content:
        - <u>User satisfaction :</u> Benefits in using EGEE, Quality of the grid, Quality of the user support, Quality of the documentation
        - <u>Parameters for survey analysis :</u> User identification, User and application profile

- **Format: excel time sheet for the moment**

- **Status:**
    - Present version available on AWG meeting agenda
    - Feedback from R.Barbera, J.Montagnat
    - Form sent to NA4 AWG on 04/03/05 for validation

User Survey: Roadmap

- **Survey available to the whole EGEE community via grid certificates**
  - Survey available by March 21st on NA4 web site
  - Survey collection and analysis using the same tools as the ones used for NA3 training survey
- **Roadmap**
  - Objective : Presentation of the first results in Athens **(18/04/05)**
  - Validation of User Survey **March 14th**
  - Circulation of the form in march **< March 21st**
 - Date of return **< April 11th**

- **Date for external review milestone: move it from PM12 to PM14**

- **Status of mini-MoUs: deadline for PEB comments is March 14**

- **What is the future strategy towards new applications (e.g. WHO, EGAAP meeting in November) ? New call in the summer with clear information on the services (not) made available**

- **NA4/SA1 meeting: what is the plan/status with respect to the Industry Forum to investigate commercial license mgmt use case? IF will look at it but only at a theoretical level, use case documented by NA4 on wiki site**

- **What is the migration plan for GILDA to gLite? What are the constraints and conditions? First evaluation at Catania, migration of Genius to gLite, migration of resources to gLite (need to share worker nodes with LCG2)**

- **What is the migration plan to gLite for the applications already deployed (e.g. on GILDA, Production Service) on LCG-2? What are the constraints and conditions? Validation of stability and robustness**

Medical workflow processing

- Taverna: worflow management system (from MyGRID UK e-science project) calling web services
- WS wrapper to LCG2 command line interface
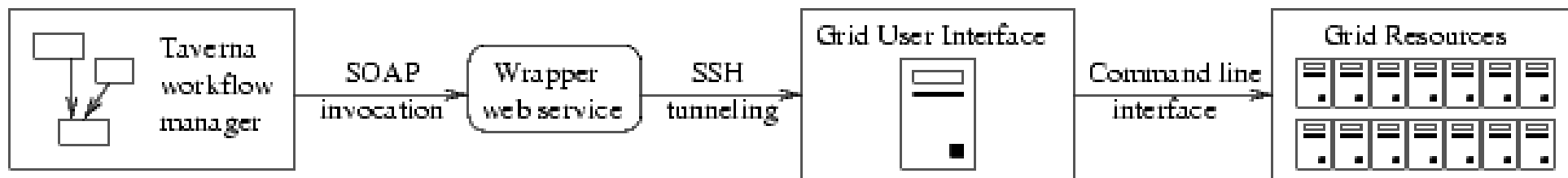
- Problems / limitations encountered
  - Synchronous nature of web services
  - Static limitations of Taverna concurrent tasks submission
  - Taking into account data dependencies (stateless nature of web services)
- Published in CBMS'05 (Computer Based Medical Systems) workshop on "Grids for medical applications"