egee

Enabling Grids for E-sciencE

Quality Assurance

Gabriel Zaquine - JRA2 Activity Manager - CS SI

EGEE 2nd EU Review

06/07th December 2005

www.eu-egee.org

Information Society
and Media

Enabling Grids for E-sciencE

- **QA objectives and organisation**
- **Overall QA achievements, metrics and main changes**
- **Plans until the end of the project**
- **Summary**

QA objectives and organisation

- **Ensure that processes, products and operational services meet the required level of quality for the project**
  - Foster the use of standards, procedures and common tools
  - Verify the project does deliver according to the agreed quality levels
    - metrics and their measurements are an important monitoring tool

- **The quality organisation has been adapted to the very large, distributed nature of the project and is composed of:**
  - JRA2 Quality Assurance Management team (QAM), 3 FTE:
    - Coordinates the overall QA for the project
    - Encourages that appropriate standards, procedures and metrics are defined within the activities
    - Ensures they are applied
    - Evaluates metrics and proposes refinements
  - Quality Assurance Group (QAG):
    - Coordinated by QAM and includes a QA representative from each activity
    - Ensures that agreed quality measures are applied inside each activity

- **Quality Group (QAG) has continued the work on refining procedures and ensuring activities can collect relevant metrics**
  - QA is in active use across all activities
- **The project has already met many of the targets set for the first 2 year phase**

Target	Current Status	End Year 2 target values
Number of Users(*)	~ 1000	≥ 3000
Number of sites	179	50
Number of CPU	~17000	9500 at month 15
Number of Disciplines	7	≥ 5
Multinational	39	≥ 15 countries

(*) Number of user certificates in VOs (excludes DTeam VO, GILDA testbed, portal users and super users)

Physics, BioMed, Chemistry, Astronomy, Earth Sciences, Geo-Physics, Banking

Current status

Other significant metrics

	Current status
Job throughput	~ 10 000 jobs/day
Percentage of sites certified as production quality	80%
Number of attendees trained	2500
Dissemination events in which EGEE-II is presented	300

Enabling Grids for E-sciencE

- **Job success rate measured during biomed WISDOM data challenge (Summer 2005)**

- **Job success rate measured during ATLAS Rome Production (January to June 2005)**

Successful jobs without WISDOM failures	70%
Failures due to the Grid/resources:	30%
Mismatching resources	12%
Failures detected by WMS	8%
Files transfer	5%
Unclassified	5%
Wisdom failures:	20%
Server licence	15%
Process errors: job distribution, human error	3,6%

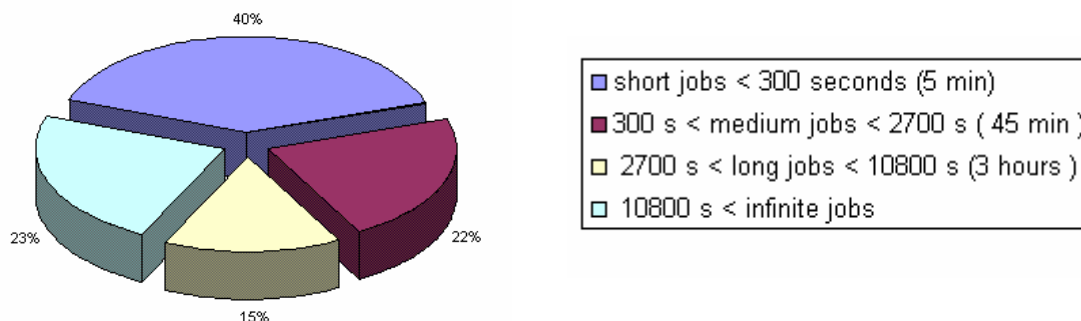|	
Successful jobs without ATLAS failures	65%
Failures due to the Grid/resources:	35%
Mismatching resources	
Failures detected by WMS	3%
Files transfer	27%
Unclassified	5%
ATLAS failures:	13%
Athena clash	9%
Proxy expired	0,3%

Enabling Grids for E-sciencE

- ## Overall Waiting time distribution

  – between UI submission and execution starting on the CE

Job type / Waiting time	%	WT
Short jobs	72	< 5mn
Medium jobs	75	< 45mn
Long jobs	81	< 3h
Infinite jobs	69	< 3h

- ## Job duration distribution (without deployment team VO)

  – between execution starting on the CE and done time stamp on the CE

- short jobs < 300 seconds (5 min)
- 300 s < medium jobs < 2700 s (45 min)
- 2700 s < long jobs < 10800 s (3 hours)
- 10800 s < infinite jobs

Enabling Grids for E-sciencE

- **Project management**
  - PPT in daily use and provided basis for period cost claims
  - Project progress, effort and cost: On track
  - Deliverables: In time delivery - at Q6: 48 deliverables + 6 Quarterly Reports + 1 Periodic report

Deliverables	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Details
Number of due deliverables	9+1	13+1	9+1+1	5+1	7+1	5+1	9+1	9+1+1	69 deliverables + 8 Quarterly Reports + 2 Periodic reports
On-time delivery of deliverables	9+1	13+1	9+1+1	5+1	7+1	5+1			Total at Q6: 48 deliverables + 6 Quarterly Reports + 1 Periodic report See details

  - We are currently defining a process and associated enhancements within PPT to better measure the performance of project partners (more than 90 partners planned in EGEE-II)

Enabling Grids for E-sciencE

Operations:

- **The program of metrics is being refined. A "control panel metrics" prototype will be available by the end of the year 2005**
  - Size metrics
    - Number of sites in production; Number of job slots; Total computing power available (kSpecInt); disk and mass storage; etc
  - Operations
    - Site response to operational tickets (open/pending/closed within timeframe); Average time of response; Site tests failed; etc
  - Usage
    - Jobs per VO (submitted, completed, failed); Data transfer per VO; CPU and storage per VO; % of sites blacklisted/whitelisted per timeframe; Number of CE/SE available to a VO
  - User support
    - Number and Time to response to user tickets (+within different groups); Number of supporters; Number of tickets escalated at various points + time related to it; etc
  - Services
    - % passing specific tests for RB, BDII, MyProxy; SRM-SE, Catalogue, VOMS, RGMA

- **Integration of monitoring information**
  - Monitoring information from various tools is collected in R-GMA archiver
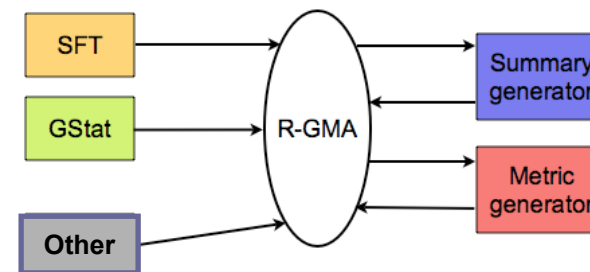
- **Applications**
  - Wider definition of "users", to include scientists who benefit from results of EGEE infrastructure, will be measured from further NA4 users' survey
  - Define a standardised way to measure the QoS seen by the applications
    - QoS provided by Grid services and Grid infrastructure
      - *Basic numbers can be retrieved via SFT (Site Functional Tests), Services' metrics and from RBs statistics*
    - Quality of the application themselves

Enabling Grids for E-sciencE

- **Continue work on procedures and major metrics and produce DJRA2.3: "*Final Report on EGEE Quality Status*"**

- **Prepare QA plans for EGEE II:**
  - Metrics not only showing quantity, but further focusing on quality of the service
  - Define a process and associated tools to better measure the performance of project partners
  - Define a standardised way to measure the QoS seen by the applications
  - Better understand the distribution for jobs submitted through RBs and other submission mechanisms

eGee

Enabling Grids for E-sciencE

- **QA is in active use across all activities**

- **We have to further focus on how to define and measure QoS**