Technical Status of the Project

Erwin Laure
(DataGrid Technical Coordinator, WP 12)

DataGrid is a project funded by the European Commission
under contract IST-2000-25182

Talk Outline

- Release 2.0 integration
- Testbed status and usage
- Post 2.0 work
- Application Working Group
- Architecture Group
- Workplan up to the end of the Project
- Summary

Release 2.0 Integration

- All software was delivered by last conference (Barcelona)
- Quite a few interaction problems and major bugs in delivered software
  - Problems tackled in quasi-daily phone conferences
  - Very productive and goal directed, labor intensive work
  - Almost daily development tags produced, up to 5 tags per day!
- Concurrent testing on LCG integration TB
  - Very helpful in quickly detecting (blocking) bugs
    - **LCG bugs treated with highest priority**
  - Helped in VDT upgrade testing
  - Additional workload due to MDS integration
- Very focused way of working kept up even during vacation period

Release Testing

- PTB on June 13th defined a number of basic release acceptance tests:
  - Job-submission:
    - Blocks of 50 minimal jobs
    - 250 jobs concurrently in the broker
    - Large input/output sandboxes
  - Data management:
    - File upload/registration/MSS storages/removal from LRC
    - File registration
  - Proxy renewal
  - Matchmaking
    - Matchmaking with input files
    - Matchmaking with input files and application tags
    - File access from within a job
- Testing team reinforced to perform the tests
- Tests were used to drive the further development
- Testing difficult due to frequent upgrade on dev. TB

Release 2.0 Integration cont'd

- **LCG produced first tag for deployment on August 7th**
- **After passing all tests (**registration quite slow due to java and consistency overheads**) 2.0 tagged on August 28th**
- **Many thanks to everybody for the hard work towards reaching this important milestone!**

Release 2.0 Integration cont'd

- All of the originally planned components except VOMS and WP1 accounting were integrated
  - Glue schema
  - R-GMA
  - LRC, RMC
  - NetworkCost
  - replica manager, optimizer
  - LCAS
  - SE
  - New WMS (MPI, APIs, interactive jobs, check-pointing,...)

- LCs involved in testing on LCG cert TB (July-August)
  - Main difference to EDG 2.0 is usage of MDS vs. R-GMA

RB Stress Tests by Massive Job Submisison

- RB never crashed
- ran without problems at a load of 8.0 for several days in a row 20 streams with 100 jobs each (typical error rate ~ 2 % still present)
- RB stress test in a job storm of 50 streams , 20 jobs each :
  - 50% of the streams ran out of connections between UI and RB. (configuration parameter – but machine constraints)
  - Remaining 50% streams finished normal (2% error rate)
  - Time between job-submit and return of the command (acceptance by the RB) is 3.5 seconds. (independent of number of streams)
- NOTE:

  RB interrogates all suitable CE's : wide area delay-killer (interactive work) ?

Preliminary full simulation and reconstruction tests with ALICE

- Aliroot 3.09.06 (including HBT correl.) fully reconstructed events
- CPU-intensive, RAM-demanding (up to 600MB ,160MB average) ,long lasting jobs (average 14 hours)
- Outcome:
  - > 95 % successful job submisison, execution and output retrieval in a lightly loaded GRID environment
  - ~ 95 % success (first estimate) in a highly job-populated testbed with concurrent job submission and execution (2 streams of 50 AliRoot jobs and concurrent 5 streams of 200 middle-size jobs)
  - MyProxy renewal succesfully exploited [OK]

Conclusions

- Some further possible tests can be performed:
  - Further investigation on DMS functionality/commands
  - Further tracking of (non show stoppers) bug fixing as EDG 2.0 evolves into 2.1
- Impressive improvement on Stability w.r.t. old 1.x EDG releases and corresponding testbeds, as stress tests confirmed
- Room for further improvements, also naturally driven by porting to newer EDG releases
- As LCG-CT will expand, **Scalability** may become the next relevant issue

Testbed Status and Usage

- Application Testbed (release 1.4) up and running until end of August
- Reduced support to free resources for 2.0 integration
- Usage quite limited

- **Successful ATLAS reconstruction tests in June**
  - 5 sites involved
  - ~250 jobs
  - ~50.000 events reconstructed

Testbed Status and Usage cont'd

- Used for tutorials and demos
  - GriDis as main infrastructure proved to be very helpful
- 2 major events with testbed including CrossGrid resources:
  - 1st GGF School on Grid Computing (Vico Equense, July)
  - CSC 2003 (Krems, September)
  - Same setup will be used for IST 2003 (Milano, October)

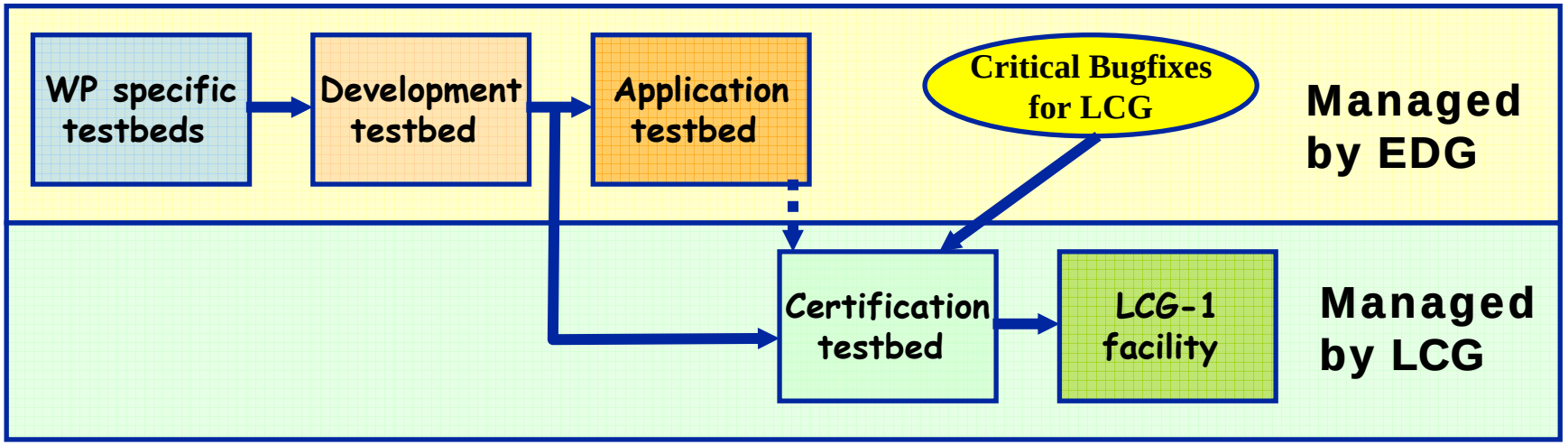
Testbed Status and Usage cont'd

- Upgrade to release 2.0 started end of August
- Main Services at:
  - RAL: LRCs, IC
  - NIKHEF: broker, LRCs
  - More services (broker) to be set up at other sites
- Currently 12 sites in 6 countries, offering ~200 CPUs
- More sites expected to join
  - Expect ~20 sites
  - Less CPU power than old testbed (LCG-1 is expected to be the production facility!)
- **Open for LC and application evaluation!**

DataGrid / LCG interaction

- EDG 2.0 components are being integrated in LCG-1 service
  - Resource Broker
  - Data management tools
  - Gatekeeper
  - Fabric and Virtual Organisation management

Interaction of EDG / LCG testbed / services



Post 2.0 Work

- PTB on August 26th decided on priorities:
  - gcc 3.2.2 upgrade (acc. September 12th)
  - VOMS server setup (acc. September 19th)
    - Initial testing for ITeam and WP6 VOs
  - VOMS integrated with job-submission
    - LCAS & LCMAS (acc. September 19th/24th)
    - WP1 using VOMS cert. (only on INFN TB yet)
  - Java security framework
    - Deployed, but not fully configured (acc. September 25th)
  - Full RLS (LRC,RLI) (acc. September 25th)
  - Secure SE setup (currently running both, insecure and secure)

- MW WP produced many other things
  - Will be demonstrated on private WP testbeds

Post 2.0 Work cont'd

- Many bugfixes:
  - WP1, WP2, WP5
  - WP3 re-factored streaming to be more efficient and less resource hungry – being tested

Impressive progress during the last three weeks

Congratulations to everybody involved!

- After the development release passes the basic tests defined for 2.0 it will propagate to application TB.
- Try to minimize interventions on application TB to avoid interference with application evaluation.

Release Experiences

- Quite limited (only 3 weeks of integration work after 2.0)
- Bugzilla used extensively – very helpful
- Autobuild and CVS used throughout all the modules
  - Dependency and rpm problems greatly reduced
- Internal testing within WPs improved significantly
  - Still room for improvements
- QAG produced a checklist
  - Initial feedback very positive
  - Helping in driving the release process

AWG

- Representatives of the three applications working group had the following tasks
  - Joint collection of requirements from the testbed experiences
  - Identification of services relevant to a common application layer
  - Description of services relevant to a common application layer
- AWG has met several times since March 2004
  - Phone conferences in July and September
  - Meetings in Amsterdam (4/03), Barcelona (05/03) and here…

AWG results

- Joint list of recommendations (06/03)
  - Goal : provide feedback to the middleware workpackages from the common needs and priority requirements of all applications
  - Available on EDMS
- Joint list of use cases
  - Goal : identify common and specific use cases for LHC, Earth Observation and biomedical applications in a grid environment
  - Available on EDMS for Heidelberg
- Enhanced working relationships
  - Forum for discussing common problems

AWG : future activities

- Contribute to prepare take off of EGEE applications
- Report on the joint testing of EDG2.0/2.1 middleware
- Evaluate external packages for high level interface to the grid (ex. Grid Application Toolkit from Gridlab project) in cooperation with other projects
- Continuing liaison with LCG work in this area (ARDA + GAG)

ATF

- Continued to review architecture
  - Post 2.0 components
- Baseline API document produced
  - Input for follow up projects
  - http://eu-datagrid.web.cern.ch/eu-datagrid/ATF/BaselineAPI/baselineapi.html
- Active in GLUE development
  - Sergio Andreozzi joined ATF as Glue consultant
- Final architecture document in preparation
  - Should be finished by final review

Workplan up to End of Project

- Application evaluation on application TB
- Maintenance and application support
- Preparation of final deliverables

- Issues:
  - Support for application TB up to review (and probably further on)
  - Support for dev. TB (needed to test bugfixes)
    - CERN discontinued due to support and security issues
    - 3 sites: RAL, NIKHEF, LAL
  - Ensure manpower for remaining work up to review!

Schedule up to [illegible]ect

Final Testbed (Testbed 3)

September

EDG 2.0 release deployed

D1.6,2.5,3.5,4.5,5.5,6.7 sw and doc.

CD needs to be sent to EU in October

Final project conference in Heidelberg (Fr.26th – Wed.1st October)

December

D11.7, D11.9

D1.7,2.6,3.6,4.6,5.6,6.8,7.7

D8.4,9.5,10.4

D12.19

PTB planned on December 15th

Review process need to start last week of November

Need moderators and reviewers!

End February 2004

Final project review

Expected Results of Conference

- End-Report to plenary should include:
  - Support structures and plans
    - Code maintenance
    - Application support
  - Plans for final evaluation
  - Plans for final deliverables