

WP10 Biomedical Applications

Final Project evaluation of EDG middleware, and
summary of workpackage achievements

V. Breton (CNRS)

Email:

breton@clermont.in2p3.fr



Outline

- ◆ Overview of objectives and achievements
- ◆ Lessons learned
- ◆ Exploitation plan
- ◆ Concluding comments
- ◆ Questions and discussion

Objectives



- ◆ To demonstrate the relevance of grids for life science
- ◆ To test the EDG middleware and feedback requirements to the middleware developers
- ◆ To raise awareness on the impact of grids in the life science community

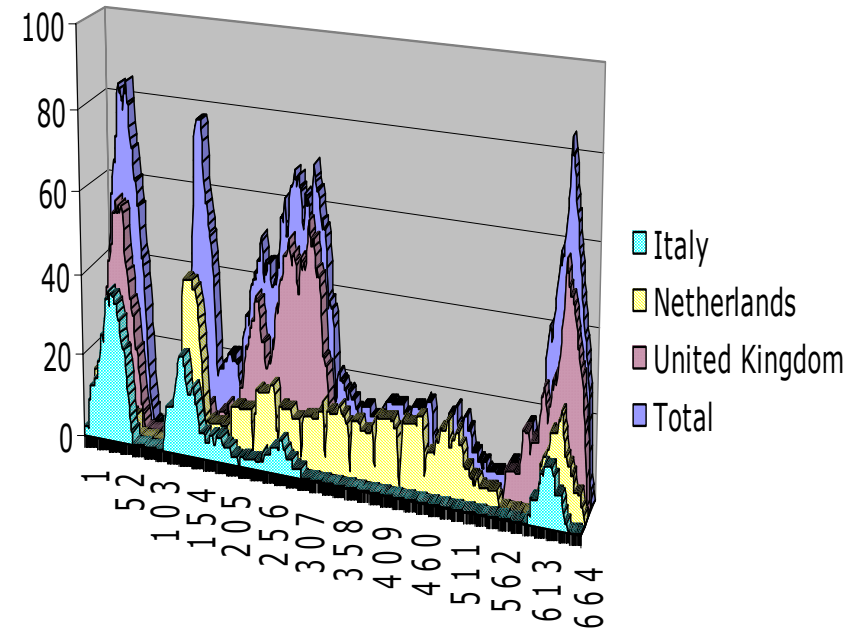
Demonstrate the relevance of grids for life science



- ◆ When DataGrid started, no clear view of where grids should apply to life sciences
- ◆ First year dedicated to identify potential applications of grids to life science
 - WP focus enlarged from biology to biomedical sciences (medical imaging)
 - Deliverables D10.1 on requirements and D10.2 on applications identified for deployment
 - First demo on distributed computing for a bioinformatics web portal
- ◆ Second year dedicated to the applications deployment on EDG testbed 1 and testbed2 (EDG1.4.x)
 - 4 out of 11 applications successfully deployed by March 2003
 - Second demo on medical images handling in a grid environment
 - Deliverable D10.3 (March 2003)

Demonstrate the relevance of grids for life science (II)

- ◆ Third year dedicated to the applications deployment on EDG testbed 2 and testbed 3 (since October 20st, 2003)
- ◆ Large scale deployment of biomedical applications was successfully achieved
- ◆ Results submitted to several journals of computer science (Parallel Processing Letters, Method of Information in Medecine,...)



Successful deployment on EDG testbed2 of large scale phylogenetics analysis (450 jobs)

DataGrid : status of biomedical applications



◆ Bio-informatics

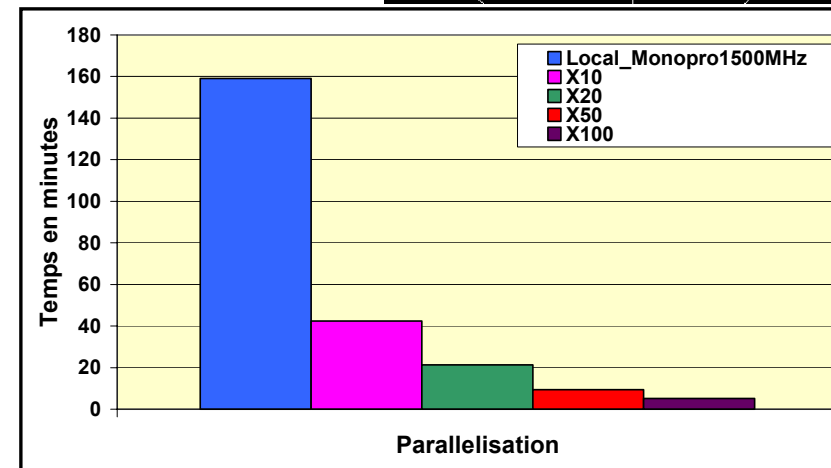
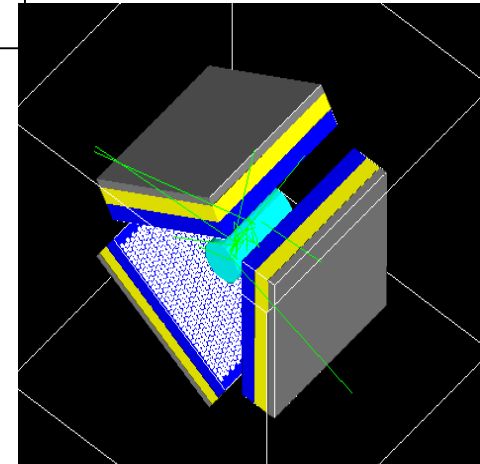
- Phylogenetics : BBE Lyon (T. Sylvestre)
- Search for primers : Centrale Paris (K. Kurata)
- Bio-informatics web portal : IBCP (C. Blanchet)
- Parasitology : LBP Clermont, Univ B. Pascal (N. Jacq)
- DNA chips analysis portal : Karolinska (R. Martinez)
- Geometrical protein comparison : Univ. Padova (C. Ferrari)

- **deployed**
- **tested on EDG**
- **under preparation**

◆ Medical imaging

- MR image simulation : CREATIS (H. Benoit-Cattin)
- Medical data and metadata management : CREATIS (J. Montagnat)
- Mammographies analysis ERIC/Lyon 2 (S. Miguet, T. Tweed)
- Simulation platform for PET/SPECT based on Geant4 : GATE collaboration (L. Maigne)

GATE Monte-Carlo simulation platform for nuclear medicine



Achievements : test EDG middleware and feedback requirements

- ◆ Requirements and testbeds evaluation described in several documents :
 - June 2001 : deliverable D10.1 on biomedical requirements
 - March 2003 : deliverable D10.3 on our experience of testbed 2 (EDG1.4)
 - October 2003 : contribution to the joint list of use cases (WP8, WP9, WP10) edited by the Application Working Group
 - November 2003 : deliverable D10.4 on our experience of testbed 3 (EDG2.0, 2.1)

- ◆ Requirements beyond HEP needs can be summarized as follows :
 - Fine grain Access Control List to files on grid Storage Elements
 - Ability to submit parallel jobs (MPI on grid clusters)

Raise awareness on the impact of grids in the life science community

- ◆ When project started, life science community showed skepticism and mistrust
 - “physicist” project
- ◆ WP10 meetings were widely open right from the beginning
 - Invited talks from EBI, EMBnet representatives
- ◆ A real momentum was gained during year 2002
 - Participation to the project of a grid for bioinformatics (EBI)
 - Up to 34 users coming from 15 laboratories in 4 different countries (April 2003)
 - Many users lost because of the delay in delivering a stable environment
- ◆ Year 2003 : birth of the Healthgrid initiative

The Healthgrid initiative

- ◆ Healthgrid: eInfrastructure for health (life sciences, drug discovery, healthcare,...)
 - long term vision : no single project can make it happen
- ◆ The Healthgrid initiative provides a glue between the projects
 - To foster exchange between projects, end users and technology developers
 - To avoid reinventing the wheel
 - To improve the take-up of grid technology
 - To disseminate information on grids for health
 - Summaries and links to health related grid projects
 - Available tools (software platforms, middleware,...)
 - Tutorials
 - Conferences
 - To promote standards
 - Involvement in GGF Life Science Research group
- ◆ History
 - First invitation to present DataGrid biomedical activities at a conference on the synergy between bio- and medical informatics in December 2001
 - Creation of the Healthgrid cluster of projects in September 2002
 - First and second Healthgrid conferences in Lyon (January 2003), Clermont-Ferrand (January 2004), next in Oxford (2005)

Failures/limitations

- ◆ Installation of a grid node in a biomedical laboratory was not achieved
 - Node installation and configuration were never mastered in WP10
- ◆ No sufficient resources to test thoroughly testbed3
 - New functionalities of great interest for WP10 were made available (parallel job submission, data management)
- ◆ No task force ever built between WP10 and middleware work packages

Lessons learned

- ◆ The importance of dialog among applications
 - Application Working Group experience was extremely positive
 - Joint list of use cases of direct use for the definition of EGEE middleware
- ◆ The importance of dialog with middleware
 - Impossibility to use EDG middleware on the sole basis of user guides
- ◆ The importance of providing a stable environment to attract the biomedical community
 - Large fraction of the users lost during the third year of the project because of testbed2 instability
- ◆ The importance of allocating resources close to middleware developers
 - WP10 resources allocated too close to “end”-users

Exploitation plan



◆ EGEE

- Biomedical activity within NA4 will take advantage of the experience acquired in DataGrid
- 1 or 2 applications may be selected for early deployment on EGEE infrastructure

◆ Other FP6 projects

- Euromedim2 (NOE, LifeSciHea, DG Research) : simulation for molecular imaging in a grid environment
- Embrace (NOE, LifeSciHea, DG Research) : grid for bioinformatics

◆ Healthgrid, an eInfrastructure for health

- Healthgrid white paper including WP10/AWG work on requirements and use cases

Concluding comments



- ◆ Initial objectives were reached
 - To demonstrate the relevance of grids for life science
 - To test the EDG middleware and feedback requirements to the middleware developers
 - To raise awareness on the impact of grids in the life science community (for instance the Healthgrid initiative)
- ◆ Large scale deployment of grid biomedical applications was achieved for the first time in Europe
 - Bio-informatics web portals have been partially grid-enabled
 - Large scale simulations for medical imaging and radiotherapy are routinely achieved on DataGrid testbed
- ◆ However, WP10 management has been a challenge
 - Cultural gap between middleware developers and WP10 users
 - Resources allocated too close to “end”-users
- ◆ A pioneering work has been done ...