Demonstration of the First Prototype of RUGBI, Design and Deployment of a Grid for Bioinformatics

Christophe Blanchet, Vincent Breton, G. Deléage, N. Démésy, F. Hernandez,
N. Jacq, S. Langlois, A. Lecluse, Yannick Legre, D. Linglin, et al.

Demonstration of the First Prototype of RUGBI, Design and Deployment of a Grid for Bioinformatics

C. Blanchet[a], V. Breton[b], G. Deléage[a], N. Démésy[c], F. Hernandez[d], N. Jacq[b], S. Langlois[c], A. Lecluse[a], Y. Legré[b], D. Linglin[d], J.F. Musso[c], R. Nougarede[c], H. Prévoteau[e], S. Reynaud[d], J.P. Roebuck[c]

[a]*Institut de Biologie et Chimie des Protéines, CNRS, France*
[b]*Laboratoire de Physique Corpusculaire, IN2P3 – CNRS, Univ. Blaise Pascal, France*
[c]*Communication & Systèmes, CS-SI,France*
[d]*IN2P3/CNRS Computing Centre, France*
[e]*Biopôle de Clermont Limagne, France*

Abstract. RUGBI is an industrial and academic project to design and deploy on top of existing technologies a computing grid offering a set of grid and bioinformatics services to analyse proteins. It aims to support life sciences SMEs for computing and storage, to deploy an interregional grid for bioinformatics and to create a biologists community in a grid environment. The proposed demonstration presents the first prototype of RUGBI architecture and bioinformatics services.

Keywords. RUGBI grid, protein analysis, grid services, bioinformatics services

Since the last decade, the technological developments for the computing grids were focused on storage, computing and resources management in order to create a very powerful environment for the management and the analysis of scientific data. Therefore, the aim of a grid is to place at the disposal of the users a geographically distributed collaborative environment which enables them to achieve more efficiently their objectives.

RUGBI [1] is an industrial and academic project funded by the French Ministry of Research (Réseau GenHomme Action Génomique et Innovations Médicales 2002 [2]). This computing and data grid is designed and deployed on top of existing technologies to offer a set of services to analyse proteins. It aims to support life sciences SMEs for computing and storage, to deploy an interregional grid for bioinformatics and to create a biologists community of biologists in a grid environment. The requirements expressed by three companies [3] were collected to make sure the RUGBI grid proposes solutions to real needs.

The demonstration presents the first prototype of the RUGBI grid. RUGBI bioinformatics facilities provide access to automatically updated genomics and proteomics databases as EMBL [4], Swissprot [5] and PDB [6], bioinformatics software as Blast [7], metabolic pathway analysis and protein secondary structure prediction tools available on the web portal SecProt [8], and *in silico* experiment management tools, through a user-friendly portal in a secure, confidential-aware and industrial environment. Performance, security and transparency of the web portal are enhanced by the distributed, interoperable and close to the portal grid information system (workflow, access controls…). Based on existing grid technology such as Globus Toolkit 3.2 and OGSI, reusable and security-aware (confidentiality and integrity) grid middleware components [9] has been developed within the framework of this project. Innovative components for managing the coordination of various tasks types (jobs, transfers of data sets, service operation invocation, …) and for aggregating, structuring and filtering log events, are examples of services developed in this project.

References

[1]   http://rugbi.in2p3.fr
[2]   http://www.genhomme.org/english
[3]   Avidis (http://www.avidis.fr); Metabolic Explorer (http://www.metabolic-explorer.com); Greentech (http://www.greentech.net/index_uk.html).
[4]   http://www.ebi.ac.uk/embl/index.html
[5]   http://us.expasy.org/spro
[6]   http://www.rcsb.org/pdb
[7]   http://www.ncbi.nlm.nih.gov/blast
[8]   http://secprot.ibcp.fr
[9]   http://grid.in2p3.fr/index.php?id=doc_rugbi