Chapter XXIX
The LIBI Grid Platform for Bioinformatics

The LIBI Grid Platform Developers, Italy

Copyright © 2009, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

ABSTRACT

The LIBI project (International Laboratory of BioInformatics), which started in 2005 and will end in 2009, was initiated with the aim of setting up an advanced bioinformatics and computational biology laboratory, focusing on basic and applied research in modern biology and biotechnologies. One of the goals of this project has been the development of a Grid Problem Solving Environment, built on top of EGEE, DEISA and SPACI infrastructures, to allow the submission and monitoring of jobs mapped to complex experiments in bioinformatics. In this work we describe the architecture of this environment and describe several case studies and related results which have been obtained using it.

INTRODUCTION

Modern biology, and particularly the molecular aspects of genomics, proteomics and transcriptomics, require multidisciplinary skills and the use of high-level technologies. Bioinformatics has been developed to meet these needs; it encompasses biological knowledge and expertise in advanced information technology. Public and private interests and investments have definitely boosted this discipline in many European and non-European countries. In Italy in particular many important initiatives have been funded to accelerate the development of bioinformatics. Among them, a project named LIBI (www.libi.it) (International Laboratory of BioInformatics, starting date September 2005), aims at setting up an advanced bioinformatics and computational biology laboratory, focusing on basic and applied research in modern biology and biotechnologies.
LIBI has its headquarters in Bari, Italy. It has been conceived as a bioinformatics laboratory “without walls” and carries out a number of activities, including: i) multidisciplinary collaborations to encourage interactions between researchers in public institutions and industry working in the biomedical and biotechnological field; ii) technological transfer and dissemination to promote training and provision of bioinformatic services; iii) the implementation of a high performance distributed infrastructure supporting the access to large data sets and high computational power; iv) the promotion of bioinformatics research at the (inter)national level maintaining and creating new specialized databases, developing new algorithms, developing and implementing new and more efficient processes and mechanisms for automatic bioinformatics analysis.

Even though this project features only Italian partners the idea is to open the laboratory to other international research groups in order to test and therefore improve the offered services.

To achieve a high level of transparency we designed and developed the LIBI Grid portal which represents the entry point to the LIBI platform, allowing easy and persistent access to distributed and heterogeneous data and software.

The LIBI project will terminate in 2009 and for four years more it will guarantee a virtual working space for the academic and industrial partners belonging to public and private institutions which shall provide their infrastructures, use the platform tools and equipment, and benefit from the same capacity and expertise made available by the other partners. The activities, technological platforms and applications will be integrated within the same environment, aiming at producing and increasing scientific knowledge, providing services and training as well as enhancing international competitiveness.

In this chapter a description of the project will be presented, proposing the architecture of the Grid PSE and the main services grouped in computational and data grids; finally, several case studies implemented with this infrastructure will be discussed as well.

BACKGROUND

A key requirement, considered during the design phase of the LIBI laboratory, has been the consideration that bioinformatics applications are naturally distributed, because experimental data and biological databases are themselves usually distributed. Also many experiments require huge computing power, owing to the large size of data sets and the complexity of processing, and may need to access heterogeneous data, where heterogeneity is multifaceted (data format, access policy, distribution, etc.), and require a secure infrastructure to protect and secure the access to private data owned by different organizations.

The Problem Solving Environment (PSE) (Houstis, 1997) is an approach and a technology that can fulfill such bioinformatics requirements. The PSE can be used for the definition and composition of complex applications, hiding programming and configuration details to the user that can concentrate instead only on the specific biological problem. Moreover, computational grids (Foster, 1999) can be used for building geographically distributed collaborative problem solving environments and grid-aware PSEs (Laszewski, 2001) can search and use dispersed high performance computing, networking and data resources. In this work, the PSE solution has been chosen as the integration platform for bioinformatic tools and data sources.

THE LIBI - INTERNATIONAL LAB FOR BIOINFORMATICS

The main goal of this project is to set up and test an advanced bioinformatics and computational biology laboratory, focusing mainly on the following activities: