Contribution ID: **175** Type: **oral presentation**

Alternative Splicing Prediction on EGEE infrastructure

Thursday 10 May 2007 14:40 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

Alternative splicing (AS) is increasingly emerging as a major mechanism in the expansion of transcript and protein complexity in eukaryotes. Computational resources for AS prediction available on the web are limited to the analysis of single genes and their related transcripts. Among these class of algorithms we are planning to use ASPic (Alternative Splicing Prediction) resource on the EGEE infrastructure.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

We are planning to distribute the large amount of task that this application requires in order to reduces the total running time; we must also solve problems like dependencies of the each task and data transfer for both input and output files. We must address also checking of failures and retries in order to make possible the running of the entire analysis as a Grid jobs.
The most important services needed in order to run this application is the WMS in order to submit jobs optimizing the utilization of each resource available, taking care of using the ranking that allow the job to run faster.
Another important service is the Data Management. We will exploit both data transport using gsiftp and LFC (LCG File Catalog) in order to simplify the storing and reading of the files.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

It seems that the services provided are enough to allow the proficient run of many distributed instance of these applications.
In the future we will provide also a web interface in order to make this service available to all users that do not know the details of the EGEE infrastructure.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

The use of bioinformatics applications on the grid allows the mining and alignment of large amount of sequences. In order to analyse transcriptome and proteome complexity of multicellular organisms, we are facing the problem of determining the alternative splicing pattern of a huge list of genes and their collection of related transcribed sequences. We are planning to do test experiments on the EGEE grid performing thousands of relatively small independent tasks, each of which costs at most minutes or hours, to analyze gene classes involved in human health and disease. This could provide increasingly important results in many areas of basic and applied biomedical research.

Author: Dr DONVITO, Giacinto (INFN-BARI)

Co-authors: Dr MIGNONE, Flavio (Università di Milano); Prof. MAGGI, Giorgio Pietro (INFN-BARI + Politecnico di Bari); Prof. PESOLE, Graziano (Università di Bari); Dr D'ANTONIO, Mattia (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita' e Ricerca); Dr BONIZZONI, Paola (DISCo, University of Milan Bicocca); Dr D'ONORIO DE MEO, Paolo (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita' e Ricerca); Dr RIZZI, Raffaella (DISCO, University of Milan Bicocca); Dr CASTRIGNANO', Tiziana (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita' e Ricerca)

Presenter: Dr DONVITO, Giacinto (INFN-BARI)

Session Classification: Experience with application domains