EGEE User Forum

Contribution ID: **150** Type: **poster**

Grid Computing for Phylogeny and Molecular Databases

Wednesday 9 May 2007 17:30 (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).

At the Institute of Zoology and the Institute of Ecology at the University of
Innsbruck molecular studies play a crucial role in scientific work. Common aim of our
cooperation is the usage of grid infrastructure for molecular biology applications:
We want to 1) apply BLAST software (Altschul et al, Basic Local Alignment Search
Tool, J. Mol. Biol. 215:403-410 (1990)) and 2) introduce software for the
reconstruction of phylogenetic trees.

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 tested WU-BLAST 2.0 (Washington University BLAST) with larger numbers of sequences
on AustrianGrid infrastructure and plan to run WU-BLAST 2.0 with large databases on
the EGEE grid.
Concerning reconstruction of phylogenetic trees two applications are now tested and
prepared for use on grid infrastructure: TREE-PUZZLE 5.2 and software of the ARB
project (Technical University Munich, Germany). Both routines can handle large data
sets of molecular sequences.

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)

As molecular biology is very fast evolving the demand of high throughput computing
will increase in the next years. The more molecular sequences are known the more
detailed analyses can be taken but the more time and computing power will be
demanded. Therefore it is necessary for molecular biology groups to have access to a

grid infrastructure to be able to perform statistical analyses and phylogenetic reconstructions.

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 usage of the EGEE grid infrastructure for biological applications will ease the usage of molecular data. 1) BLAST software allows a rapid sequence comparison and directly approximates local similarity. Regularly large scale comparisons will facilitate finding genes of interest in the vast number of –for instance - EST sequences (expressed sequence tag). 2) Molecular sequence data can be reconstructed by use of statistical analyses like maximum likelihood or pair wise maximum likelihood distances. Depending on the number and the length of amino acid sequences this reconstruction demands enormous computing power.

Author: Dr NIMETH, Katharina (University Innsbruck)

Presenter: Dr NIMETH, Katharina (University Innsbruck)

Session Classification: Poster and Demo Session

Track Classification: Poster session