



Contribution ID: 111

Type: **Poster**

A Grid-Enabled Problem Solving Environment for QTL Analysis in R

Monday, April 12, 2010 6:30 PM (3 minutes)

There exist a large number of applications in multidisciplinary research environments that require efficient, seamless, and easy-to-use infrastructures to address computationally intensive problems. In this work we present an approach to build Grid-enabled problem solving environments (PSE), that allow end-users to operate within familiar settings and provide transparent access to computational Grid resources.

Conclusions and Future Work

The proof-of-concept implementation has illustrated that this type of system can be implemented with relative flexibility using pre-existing components. The PSE integrated into R can be further enhanced with existing code packages and tools, given proper interfaces and data translation services.

Detailed analysis

Quantitative Trait Loci (QTL) analysis is identified as a suitable candidate for exploitation of large-scale computing infrastructures such as Grids. Our approach is illustrated in a proof-of-concept implementation of a PSE for multidimensional QTL analysis. The goal of such analysis is to determine a set of loci in the genome that are responsible for the genetic component of a quantitative trait, using phenotypic and genotypic data gathered from an experimental population. The analysis uses a powerful statistical model, and the computational requirements rapidly become very large when several loci are searched for simultaneously.

The prototype integrates a system built on the open source statistical software system R with the Grid Job Management Framework (GJMF) which provides middleware-transparent access to Grid functionality, allowing our PSE to not be limited to using a specific middleware, but rather offload Grid issues to existing arbitration layers.

Impact

The QTL PSE is targeted towards end-users with limited experience of high-performance computing, and supports workflows where small tasks are performed locally on PSE hosts, while larger, more computationally intensive tasks are allocated to Grid resources. In this model, Grid computations are scheduled asynchronously, allowing client hosts to perform e.g. visualization and postprocessing of data in parallel with large-scale Grid computations. The proposed integration model provides the end-users with all the capabilities of the R environment coupled with transparent access to the computational power of grid environments. Use of the R-based environment allows end-users to operate in environments they are familiar with, and provides statistical processing capabilities. Use of the GJMF decouples the PSE from specific Grid middlewares and provides abstractive and reliable access to Grid resources through concurrent use of multiple Grid middlewares.

Keywords

QTL, workflows, PSE, R, bioinformatics, middleware arbitration

URL for further information

<http://www.it.uu.se/research/project/ctrain>, <http://www.gird.se/gjmf>

Primary authors: Mr NETTELBLAD, Carl (Department of Information Technology, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden); Mr JAYAWARDENA, Mahen (Department of Information Technology, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden); Mr ÖSTBERG, Per-Olov (Department of Computing Science and HPC2N, Umeå University, SE-901 87, Umeå, Sweden); Mr TOOR, Salman Zubair (Department of Information Technology, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden)

Co-authors: Dr ELMORTH, Erik (Department of Computing Science and HPC2N, Umeå University, SE-901 87, Umeå, Sweden); Dr HOLMGREN, Sverker (Department of Information Technology, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden)

Presenter: Mr TOOR, Salman Zubair (Department of Information Technology, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden)

Session Classification: Poster session

Track Classification: End-user environments, scientific gateways and portal technologies