Minutes of the VC meeting on 'The Life Sciences Virtual Community '

Moderator – Mr. Andrew Lynn, JNU, Delhi
Date and Time – 04 Mar 2008 - 12 PM to 12:40 PM

While implementing grid-middleware is an essential part of the establishment of grid infrastructure, data-sharing, and collaboration are relatively independent of the successful implementation of the grid middleware which will also provide a uniform-view of compute nodes. This has been largely ignored by the Life Sciences community, and we would like to make a correction with the following mini-projects

Data Availability: There is a need to implement a biological data-mirror, on the lines of the APAN-AP-BioNet (http://www.bio-mirror.net/). The size of the compressed databases is in the range of 800 GB, and is growing exponentially. ERNET – who are members of APAN, and JNU – maybe approached for incremental updates using their APAN link. JNU will provide storage and a portal, and either through the GARUDA link, or a separate ERNET link, sync the databases with the primary mirror, and make them available to the GARUDA community using a portal. Arrangements have been made to ship the databases initially from the US site.

Collaboration: One of the major uses of the high-speed connectivity is to encourage unlimited audio-video communication between personnel at the member institutes. Access to research personnel at member institutions would not only foster research collaborations but also provide the variety of skills useful in multidisciplinary subjects such as Computational Biology. In addition, the creation of a collaboration suite/portal for project management, data sharing and dissemination would be useful. Access-Grid technologies developed for the purpose of collaboration has also been successfully implemented, though the use of a simpler learning management system/content management system with audio-video multicasting could be used as a less expensive replacement. Collaboration at the research level using a sharing of high-performance computer resources and communication could be extended by the use of the grid for real-time linking of instrumentation facilities and laboratories for fast data interchange and telemedecine. In the proof-of-concept stage, we would like to explore the use of a point to point between SGPGI. Lucknow and RCC, Trivandrum, and multicasting with IIT Kgp which has a medical department, and JNU. In the future, the development of databases by partner institutions would be explored, through the use of Medical images – though formats, confidentiality, and copyright remain issues to be discussed.

Education: As an experiment useful for education in computational biology – it is proposed to explore the possibility of development of modules in a subject and its dissemination using EDUSAT. In addition, by drawing on the skills of a wider pool within the virtual community we can explore possibilities that arise from multidisciplinarity. We suggest the building and integration of repositories at institutions on the grid – such as the "million book project", audio lectures from the Educational

Technology projects at IITs, and course based content from users managed using a learning management system. Eventually we hope to develop this as a **data-grid** with data repositories developed at partner sites being shared.

Computational Applications: In general, it is useful to define applications as high-throughput or high-performance. Applications from the Life Sciences contain instances of both: virtual screening and molecular dynamics simulations are examples for each respectively. Each member institute on the grid already runs applications used by their in-house personnel. A semantic approach to collecting information on resources related to hardware and operating systems has already been developed. This could be extended to applications available at each member-site.

For Life-Sciences applications, we listed four tasks to be implemented during the POC phase. The following were test conditions which have been implemented.

1. Data availability

2. Development of a Grid Portal
BRAF has already been developed by Rajendra Joshi along with its applicability for the grid – Genome grid. This will serve as the primary access to applications on the grid.

3. Commonly used applications were selected to test coupling, heterogeniety of nodes, parallel libraries
(i) Embarrasingly parallel tightly coupled scale easily
BLAST – uses MPI
HMMER – uses PVM
(ii) Embarrasingly parallel loosely coupled
Autodock
(iii) Parallel tightly coupled – does not scale easily
Gromacs
NAMD-G – a grid enabled version of the popular . This was eventually not implemented as the use of the CHARM++ libraries used to parallelise the application was not standard.

(4) In-house applications
An application that uses MPI for heart simulation was selected from Institute of Mathematical Sciences, Chennai. It was successfully implemented across multiple clusters of the grid.

Summary:
Life Sciences applications have been ported and tested on Garuda. A portal for applications is being developed. These applications have been largely based on HPC applications. We propose services based on data-sharing, collaboration (communication) and education to be implemented as non-HPC uses of the grid.