Contribution ID: **130**   Type: **Demo**

Interactive Virtual Screening on the Grid

Tuesday 26 September 2006 17:00 (20 minutes)

Recent studies have suggested that the high pathogenic avian flu H5N1 virus has the
potential of developing drug resistance and of acquiring the ability of
human-to-human transmission. To enable biologists a better response to the threat,
the second EGEE biomedical data challenge battling avian flu was set to screen
300,000 compounds against 8 predicted mutations of the Influenza A Neuraminidase for
analyzing the efficiency of the known drugs and for searching new drugs. In April and
May 2006, we succeeded to mobilize over 2,000 CPUs in the EGEE Grid infrastructure,
demonstrating that the high-throughput screening (HTS) of drug analysis can be
efficiently reproduced on the Grid using the WISDOM platform previously developed for
the Malaria data challenge in last Summer. The 6-weeks activity has covered over 100
CPU years of CPU power required for the virtual screening process and has produced
about 600 Gigabytes of docking results for further analysis.

Current computing model of the Grid-enabled HTS adopts a coordinative way of
execution in order to gain the docking throughput; however, to bring biologists a
real end-user application for their daily research, the application usability needs
to be improved taking into account the realistic usage patterns. For example, the
preparation and deployment effort needed for starting the data challenge will not be
appreciated by the users who frequently repeat the virtual screening for testing
their libraries and docking parameters. The batch mode HTS is also not feasible for
interactive analysis which can save biologists'time by allowing them to start
analyzing partial results on the fly instead of dealing with a huge amount of output
at the end. In addition, biologists prefer a graphical user interface to configure
domain-specific parameters.

To improve the usability, we first introduced a light-weight framework called DIANE to enable the interactive analysis of the Grid-enabled virtual screening application. DIANE was originally developed for handling the distributed applications within a Master-Worker model. It provides an overlay system on top of the Grid system, in which the pull-mode scheduling and failure recovery mechanisms are implemented based on the CORBA protocol. On the other hand, the DIANE framework hides the details of the job operations on the Grid so that application developers can concentrate on the implementation of application logic, and end users benefit from the simplified job descriptions containing only intuitive and application specific parameters. The stability and efficiency of DIANE has been tested by taking a significant part of the avian flu data challenge.

Following the successful avian flu data challenge, the Academia Sinica Grid Computing Centre (ASGC) in Taiwan is in charge of developing a user-friendly docking environment leveraging on the DIANE framework. Building on top of the DIANE command-line interface, we have customized a web-based interface for biology end-users. Through the web interface, users can quickly set a filter on compound libraries, configure docking parameters, start-up and monitor their virtual screening activities on the EGEE production environment. During the job execution, the completed dockings are scored based on the binding energy. The docking complexes are available as soon as the results are produced; therefore biologists can progress on for further analysis without being blocked until their jobs are finished. A visualization interface of complex structures also aids biologists in analysis.

Primary author: Mr HO, Li-Yung (ASGC, Taiwan)

Co-authors: Mr LEE, Hurng-Chun (ASGC, Taiwan); Dr WU, Ying-Ta (Academia Sinica, Taiwan)

Presenter: Mr LEE, Hurng-Chun (ASGC, Taiwan)

Session Classification: Demo session

Track Classification: Users & Applications