Highlights of HPC Competition 2006: PC Grid Computing

Wang Junhong
(SVU/Academic Computing, Computer Centre)

You might have heard many success stories in recent years on the use of Grid computing technologies for bio-life science related applications, such as gene sequencing, molecular docking and drug discovery. Using the Tera-Scale Campus Grid at the National University of Singapore, or TCG@NUS in short, researchers participating in the HPC Challenge have demonstrated that grid computing resources, particularly the PC Grid, is a very powerful resource for large scale life science related applications. Significant speedup can now be achieved using the 1,200 PCs available on TCG.

Three submissions were received for the speed up challenge in the PC Grid category. The computing resource used is the TCG@NUS PC grid computing environment. It is very interesting to note that all the three projects submitted are related to life sciences, although PC grid computing is not exclusively dedicated to life science applications. Among the three projects, two projects deal directly with life science applications - one is running BLAST software to do the metagenomic analysis, the other is running HMMER codes for genome analysing. The third project studies the fluid structure for an optimal design with a rotating bioreactor using self-developed codes.

For the metagenomic analysis with BLAST, researchers can cut down the computing time tremendously by carefully splitting the big sequence task to thousands of small tasks which are then run on a few hundred computers concurrently. Similarly for the HMMER domain searches project of a fly (Drosophila melanogaster) genome, an initial successful run involved about 36,000 sub-tasks running on about 500 Windows computers can easily achieve a 4.16 times speedup. This helps the researcher shorten the computing time from an initial 17.5 days on a well configured Athlon64 3200 server to 4.2 days. With such speedup impact, researchers find that it is possible to perform the same type of analysis on larger genomes such as human and mouse, which was previously not possible.

The third project is to study of the fluid structure in a small scale bioreactor for animal cell, where a user-developed solver needs to be run many times with different input parameters according to various configurations and conditions in order to search for the optimal bioreactor design. Using TCG@NUS, the researcher is able to complete all the 100 parametric runs of the same program within 4 hours that previously needed a few days to complete on a single computer. With that kind of performance, the research clearly concludes that “the PC Grid is the best choice to save your time if you have a mass of parameter sweep jobs to process.”

Moving ahead, we will expand the TCG@NUS further to include more computer devices. The target is to have 2,000 by the end of 2006. Meanwhile, we will work closely with researchers to enable their applications running on TCG@NUS if the applications are suitable to run with PC grid computing. A web portal access for the job management will be developed for the ease of use as well by this year.

Researchers are welcome to contact us if you have interest to use TCG@NUS PC grid computing. Feel free to contact us if you have any enquiries.