

Applying the Grid Computation in Long-Term Retrospective Health Insurance Data Analysis

To the Editor—Long-term retrospective health insurance data have been a very useful objective resource in medical and health-care research [1,2]. In Taiwan, started from 10 years ago, a health information database, the National Health Insurance Research Database (NHIRD), has been created to collect complete national claim data including disease diagnosis, treatment, inpatient and outpatient claims, and details of prescription dispensed. Recently, more and more researchers try to discover and solve problems of health policy and epidemiology by analyzing NHIRD in their studies [3,4].

Data mining algorithms have been applied in retrieving useful information from unstructured databases. However, when the database gets larger and larger, it requires lots complicated data analysis procedures and computational power to do the job. In the Center of Biomedical Informatics in Taipei Medical University, serving as an NHIRD repository, we have developed a data analysis framework based on the Grid architecture, which integrates many computers as a powerful computing resource to process the NHIRD data more efficiently and robustly. The Grid was built by using 12 Mac Mini computers with 720 gigabytes (GB) storages. For the purpose of applying the Grid technology in analysis of large retrospective data, we have designed a data organizing procedure to break data off from the NHIRD and a job scheduler to manipulate the analysis processes including program allocation, job submission, error detection and correction.

Because drug interaction is a very important patient safety problem, retrospective data analysis have been used to study the severity of specific drug interactions [2]. In our ongoing study, we have been focusing on observing drug interactions in outpatient prescriptions for every clinical department in all hospitals in Taiwan from 2000

to 2002. The Grid technology makes it possible to analyze more than 600 GB NHIRD data efficiently. The drug interaction can be analyzed completely for the whole nation in one day by using our computation architecture.

Complicated problems can be simplified and improved by the help of Grid technology. Although we have developed a robust, secure, and efficient Grid-based data analysis framework for drug interaction analysis, with slight modification in the algorithms, this framework can also be applied for other long-term retrospective research topics.—Yung-Tai Yen, MS, Yu-Chuan Li, MD, PhD, Graduate Institute of Medical Sciences, and Ming-Chin Lin, MD, Chien-Yeh Hsu, PhD, Graduate Institute of Medical Informatics, Taipei Medical University, Taipei, Taiwan.

References

- 1 Subramanian S, Khandker RK, Roth D. Long-term resource use and cost of percutaneous transluminal coronary angioplasty versus stenting in the elderly: a retrospective claims data analysis. *Value Health* 2003;6:534–41.
- 2 Henriques C, Schultz D, Milanette T, et al. Using retrospective claims data to design clinical safety surveillance of thiazolidinedione drugs. *Value Health* 2001; 4:111.
- 3 Lin HC, Lee HC, Chao PZ, Wu CS. The effects of weather on the incidence of sudden sensorineural hearing loss: a 5-year population-based study. *Audiol Neurotol* 2006;11:165–71.
- 4 Shao PL, Chie WC, Wang CY, et al. Epidemiology of *Haemophilus influenzae* Type B meningitis in Taiwan, 1997 and 2000. *J Microbiol Immunol Infect* 2004; 37:164–8.