E-Science Infrastructures for Molecular Modeling and Parametrization

S. Pamidighantam*

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Illinois USA

*Author for correspondence e-mail: spamidig@ncsa.illinois.edu

In this presentation the development and deployment and operational aspects of the E-Science infrastructures of GridChem [1], a computational chemistry grid and Paramchem [2] will be described. The Computational Chemistry Grid is a virtual organization that serves computational chemistry community with a diverse set of hardware and software resources and provides services to enable access to such resources in a routine and transparent manner. The organization provides all services required for high performance computing needs for computational chemistry with allocations, application interfaces, job and data management, consulting and outreach activities. Paramchem is a similar project that provides infrastructure with dedicated client and services for molecular forcefield parametrization based on *ab initio* reference data.

The basic structure for computational services includes a three tier architecture that consists of a Client, a Web Services Layer and a HPC computational layer. The client is a Java based desktop application which consists of some tools for pre and post processing job data and some communications and data organization tools. The web services layer consists of additional layers that includes data organization and management, scientific workflow management, and communications management. This layer also provides authentication and session control protocols, resource information collections, discovery and matching, job information logging and notification and interacts with databases that holds information and authentication servers such as Myproxy [3] for credential delegation. Currently GridChem has integrated several popular applications that serve quantum chemistry, molecular dynamics and quantum *Monte Carlo* software that can utilize a distributed set of high performance systems. Tertiary storage mechanisms are integrated for long term storage of user data. Some of the services deployed are accessible directly from a web browser for wider community. GridChem serves about 400 users at present and Paramchem atom typing and initial guess services are used by more than 1000 users. The NSF funded project has been a successful science Gateway [4] under NSF-XSEDE [5] program and serves specific projects in addition to generic community service. The sustainable operation of such e-infrastructure will be discussed broadly.

References

1. R. Dooley, K. Milfeld, C. Guiang, S. Pamidighantam and G. Allen, Journal of Grid Computing, 2006, 4, 195–208.
2. J. Ghosh, S. Marru, N. Singh, K. Vanomesslaeghe, Y. Fan and S. Pamidighantam, TG '11 Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery Article No. 35
3. http://grid.ncsa.illinois.edu/myproxy/
4. N. Wilkins-Diehr, D. Gannon, G. Klimeck, S. Oster and S. Pamidighantam, Computer, 2008, 41, 32-41.
5. https://www.xsede.org/