Performance of Molecular Dynamics Application (DL_POLY_3)
20x Faster after Optimisations by HECToR dCSE Team

Ilian Todorov, Science and Technology Facilities Council (STFC)
Ian Bush, Numerical Algorithms Group Ltd (NAG)
HECToR CSE Team, Numerical Algorithms Group Ltd (NAG)

HPC experts from NAG, working under NAG's Computational Science and Engineering (CSE) support service for HECToR, the UK's national academic supercomputing facility, have improved the scalability and performance of DL_POLY_3, a widely used software package for studying molecular dynamics. The 20-fold improvement in performance achieved by this project enabled a study of egg-shell formation that was infeasible with previous performance.

DL_POLY_3 is a general purpose package for classical molecular dynamics (MD) simulations from STFC Daresbury Laboratory. University of Warwick researchers Mark Rodger and David Quigley, in collaboration with colleagues at the University of Sheffield, used DL_POLY_3 and the HECToR supercomputers to study the role of a protein called ovocleidin-17 (OC-17) in chicken eggshell formation. Significant performance improvements were needed to make the modelling possible in feasible timescales using the HECToR supercomputers (especially in terms of parallel I/O).

Commenting on the dCSE project success, David Quigley said "*Prior to these I/O improvements, DL_POLY_3 was unable to make effective use of the parallel file system on HECToR, severely crippling the performance of our simulations. The new code has reduced the time taken to write a single snapshot from 3 minutes to less than half a second, resulting in an overall factor of 20 improvement in our net performance.*"

HECToR

HECToR is managed by EPSRC on behalf of the participating Research Councils with a mission to support capability science and engineering in UK academia. The Cray XT supercomputers, located at the University of Edinburgh, are managed by UoE HPCx Ltd. The CSE Support Service is provided by NAG Ltd and ensures users have access to appropriate HPC expertise to effectively exploit advanced supercomputers for their science. A critical feature of the CSE Support Service is the distributed CSE (dCSE) programme which, through lightweight peer review, delivers dedicated performance and scalability projects on specific codes in response to proposals from users. The dCSE programme now consists of over 50 focused projects complementing the traditional HPC user applications support and training also provided by NAG.

The dCSE projects completed so far have delivered outstanding examples of the cost savings and new science that can be enabled through dedicated CSE effort. The DL_POLY_3 project reported here adds to these success stories with a successful performance improvement.

Project Background

With the increase in size (number of compute nodes) of supercomputers, it is not only the parallel performance of applications that is preventing greater exploitation - in many cases it is the I/O which is the bottleneck. This is especially the case for distributed data algorithms (as employed by DL_POLY_3). The aim of this dCSE project was to improve the I/O performance so that real science could be performed on thousands of cores when simulating large systems. This optimisation involved large scale data rearrangements to enable both good use of the Lustre file system and also an efficient parallelisation of the I/O. The modular design also allowed it to be adapted very easily to use either MPI-I/O or NetCDF.

Ilian Todorov and William Smith of STFC are the main developers of DL_POLY_3. Ilian Todorov of STFC was the Principal Investigator on the project. Ian Bush of NAG carried out the 6 person-month project in collaboration with the NAG CSE team and the DL_POLY_3 developers. It is estimated that usage of DL_POLY_3 exceeded 8 million AUs (allocation units) since late 2009 on HECToR.

Project Results

The dCSE project implemented an extensive data redistribution to allow best use of the I/O subsystem, resulting in a code that scales to many more processors, despite the large increase in communications required. These code enhancements resulted in net performance improvements of a factor of 20 for one key scientific problem (egg shell formation), with multiple order-of-magnitude performance improvements in data file reading and writing aspects of the code.

This work now allows the code to exploit efficiently thousands of cores and to simulate systems with many millions of particles, both of which were previously not possible due to the overheads incurred by I/O.

A full technical report can be found at http://www.hector.ac.uk/cse/distributedcse/reports/

For more information contact: HECToR CSE Team
The Numerical Algorithms Group Ltd, Wilkinson House, Jordan Hill Road, Oxford, OX2 8DR, United Kingdom
Telephone: 01865 511 245 Email: hector-cse@nag.co.uk Web: http://www.hector.ac.uk/cse/

HECToR – A Research Councils UK High End Computing Service