HCW Keynote: Aspects of Heterogeneous Computing in the Open MPI Environment

Richard L. Graham

Los Alamos National Laboratory
LA-UR-06-0777
NM, USA

There are several aspects to heterogeneous computing, with this talk focusing on several of these, as they relate to execution of a single parallel job. This includes processor and network heterogeneity, as well as the support for making this heterogeneity transparent in the run-time system. Design and implementation choices made by the Open MPI/Open RTE collaboration will be discussed, with an emphasis placed on the effort to make these choices transparent to the end users - both for MPI and non-MPI parallel jobs.

The main standard libraries used for scientific simulation over the last decade and a half are the Message Passing Interface (MPI) and the Parallel Virtual Machine (PVM) libraries. Early implementations of PVM supported processor and network heterogeneity, allowing a single application to run in a hybrid environment. However, this came with a significant performance penalty. Implementors of the MPI standard, the de-facto communications standard for scientific computing, have tended to focus most of their efforts on highly optimized, single system implementations, largely ignoring the challenges posed with the goal of implementing an efficient MPI for a heterogeneous environment. In addition, implementations that enable running an application in a heterogeneous environment have tended to expose these details at the application level, requiring the applications to deal explicitly with these environment, limiting the extent to which heterogeneous computing has caught on.

Building on experience gained with the LA-MPI, LAM/MPI, FT-MPI, and PACX-MPI, the Open MPI project is to enabling applications to run effectively on available hardware. This effort is aimed at removing the practical requirement restricting single application runs to a single type of hardware, and increasing the ability to utilize available hardware. Special attention is given to hiding these details from the end-user, so that, from an applications perspective, running on a homogeneous system is essentially the same as running on a homogenous system. Support is included for high performance communications in a hybrid environment, as well as run-time support for heterogeneous environments.

This talk will describe the component architecture used by the Open MPI project which forms the foundation for providing instance specific implementations of a particular functionality, such as point-to-point communications between two specific end-points. The design choice to provide fine level control over the algorithms deployed within a single job provides a key architectural feature enabling optimal use of system resources. This talk focuses on point-to-point communications and the run-time (Open RTE) environment used to create, monitor, and terminate parallel job execution - both MPI and non-MPI jobs. The point-to-point communications discussion will focus on the architectural features enabling pair-wise communications tailored to the requirements posed only by the specific pair, as well as those that enable the simultaneous use of different network types between a given pair of communication end points. Performance data will also be presented. Aspects of Open RTE that enable distributed process monitoring and control in a heterogeneous environment will be discussed.

Speaker biography: Richard Graham is the Computer Systems and Software Environment (ASC) Program manager, and the Advanced Computing Laboratory acting group leader at the Los Alamos National Laboratory. He joined LANL's Advanced Computing Laboratory (ACL) as a technical staff member in 1999. As team leader for the Resilient Technologies Team he started the LA-MPI project, and is one of the founders of the Open MPI collaboration. Prior to joining the ACL, he spent seven years working at Cray Research and SGI.

Rich obtained his PhD in Theoretical Chemistry from Texas A&M University in 1990 and did post-doctoral work at the James Franck Institute of the University of Chicago. His BS in chemistry was from Seattle Pacific University.

1-4244-0054-6/06/$20.00