

MPI Scaling Using Intel and MVAPICH2

Science Mission Directorate

This work demonstrates some of the complexities that arise when trying to develop software that is meant to run on large-scale supercomputers. It helps the Earth science and high-performance computing communities by identifying system design bottlenecks, as well as performance limitations due to application design.

We undertook a scaling and performance study comparing two implementations of the Message Passing Interface (MPI) software commonly used on high-performance computing (HPC) systems: MVAPICH2 and Intel MPI. The study included running a benchmark developed at the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. The benchmark is representative of the finite-volume cubed-sphere (FVCS) dynamical core, which is part of the GEOS-5 climate modeling system.

This study was designed to evaluate the scalability of the Discover cluster and the FVCS benchmark, while at the same time identifying any performance variability between the MVAPICH2 and Intel MPI implementations.

Results indicate that our benchmark scales very well in a clustered environment. Additionally, we were able to successfully quantify the performance overhead associated with the computational work of the benchmark versus the overhead imposed by each MPI implementation. The benchmark results show that there are indeed scalability problems in Intel's current MPI implementation, with regards to the startup and shutdown of processes, when using a large number of cores (over 1,500).

*Tyler Simon, NASA Goddard Space Flight Center
tyler.simon@nasa.gov*

Communication patterns for 90-megabyte (MB) messages of the finite-volume cubed-sphere benchmark on 1,176 processors. *Tyler Simon, NASA/Goddard*

