Open Fabrics Workshop – MPI Panel Intel® MPI

3/24/09

Bob Woodruff/Alexander Supalov

Agenda

- Intel® MPI product status
  - Intel MPI 3.2 & 3.2.1
- Intel MPI requirement of Open Fabrics
  - What we would like to see

Intel® MPI product status

- Intel MPI 3.2 & 3.2 Update 1
  - Based on ANL* MPICH2 and OSU* MVAPICH2*
  - MPI 2.0 compliant
- **Hardened commercial grade** MPI
  - In use with more than 56 ISV applications
- Available for both Linux* and Microsoft® Windows*
  - Supports all the major Linux* distributions
    - Red Hat, SLES, Fedora Core, CentOS, SGI* Propack …
  - Supports Microsoft Windows XP*, Vista*, Server 2003, 2008, and HPC 2003, 2008
- Also ships as part of a **full cluster toolkit**
  - Intel Trace Analyzer/Collector
  - Intel Math Kernel Library
  - Intel MPI benchmarks
  - Intel Compilers
  - HPCWire Editor’s Choice Award in November 2008

*Other names and brands may be claimed as the property of others.

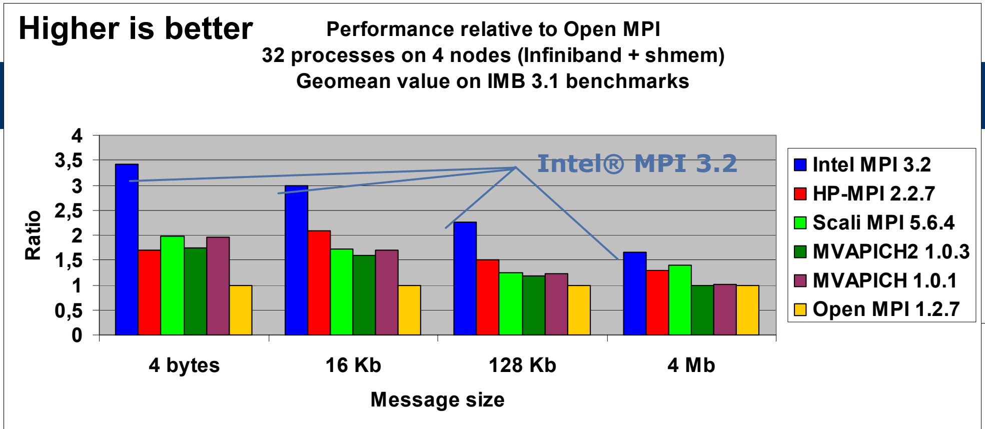
Intel® MPI product status

- Multi-Fabric support without recompilation
  - Uses uDAPL* for its stable, versioned, extensible interface
    - **Achieves outstanding performance while maintaining forward compatibility for new H/W for our ISVs**
  - InfiniBand* Architecture and iWarp support via Open Fabrics uDAPL*
  - Myrinet*, Quadrix* via stand alone uDAPL* providers
- Intel MPI 3.2 available since October 2008
- Intel MPI 3.2 Update 1 shipping this week
- New in the 3.2 & 3.2 Update 1 releases
  - Performance enhancements
  - Usability improvements
  - Extended interoperability
  - See backup for details

*Other names and brands may be claimed as the property of others.

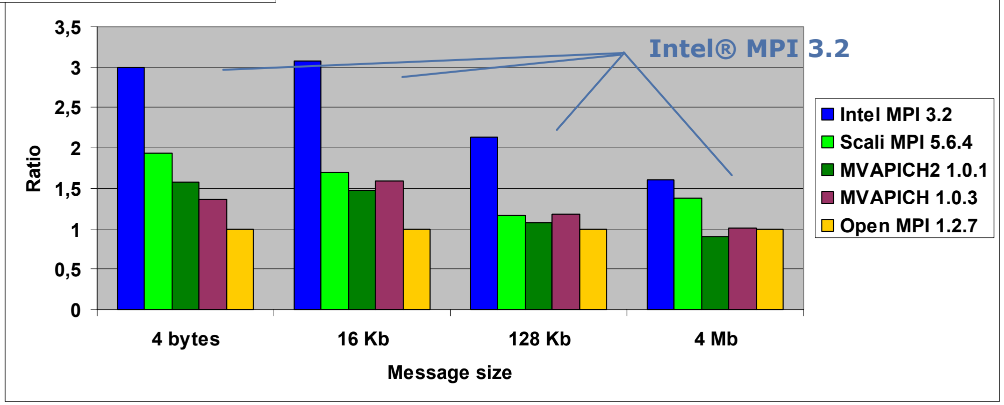
Industry Leading Performance

Intel MPI, HP-MPI, ScaliMPI, MVAPICH[2] vs OpenMPI



Performance relative to Open MPI
64 processes on 8 nodes (Infiniband + shmem)
Geomean value on IMB 3.1 benchmarks

Higher is better



Intel MPI 3.2 provides an industry leading out-of-box performance

- Incremental optimizations
- Best default parameters
- Best collective algorithms

Hardware Configuration:
Interconnect: InfiniBand, ConnectX adapters
CPU: Xeon DP Harpertown X5472 FC-LGA6 3.00Ghz 1600FSB
12M 64bit 120W (80574KL080NT)
RAM: 16Gb per system
Software Configuration:
Intel® MPI Benchmarks 3.1 *(1)*

(1) Source: Intel Corporation. Test results aggregated with overall performance scores based on geometric mean. Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference www.intel.com/software/products

Intel® MPI, Intel Cluster Ready, and Open Fabrics QA synergy

- Intel Cluster Ready
  - Clustering made simple
- Intel MPI and OFED are key ingredients to the Intel Cluster Ready Program
  - As new recipes are developed, they are tested with OFED and Intel MPI
  - As new OFED code is developed, we test it with Intel MPI and ICR recipes

Intel® MPI requests to OFA

- ABI stability
  - Need stable, versioned ABI on Linux* and Microsoft® Windows*
- Common Interfaces between Linux* and Microsoft Windows*
  - uDAPL* provides this now
  - Can we have a common I/F at the verbs level ?
- Scalability
  - More scalable, more stable, more memory efficient connection establishment mechanism for RC based services
  - Connectionless services (uDAPL* UD, etc.), also for iWarp, on Linux* and Microsoft Windows

*Other names and brands may be claimed as the property of others.

Intel® MPI requests to OFA (cont.)

- Usability
  - Transparent, efficient, cached memory registration in HW
    - If impossible in HW, OFED provided kernel hooks for memory allocation/release signaling
  - Proper interaction with system(3) and other fork(3) based calls
- Testing of Intel MPI in the IWG OFA interoperability events
  - Important for our customers to know that Intel® MPI has been tested with OFED and WinOF

Thank You

New in Release 3.2

- Performance and Scalability enhancements
  - Automatic application-specific performance tuning through the mpitune utility
  - Simplified selection of IPoIB for sock and ssm communication through the I_MPI_NETMASK variable
  - Faster process startup thanks to disabled Python* compatibility check
  - Faster RDMA and RDSSM wait mode through the I_MPI_RDMA_WRITE_IMM variable
  - Further optimized Alltoall, Alltoallv, Allreduce, Gather, Scatter, and Bcast collective operations
  - Greater scalability for the sock and ssm devices
  - Greater scalability for InfiniBand using the uDAPL* socket CM provider

*Other names and brands may be claimed as the property of others.

New in release 3.2 (cont.)

- Usability improvements
  - Advanced shared memory segment size control
  - Flexible OS, Python*, compiler, and DAPL* compatibility check control
  - Loadable 3rd party process manager libraries
- Extended interoperability
  - Intel® Compiler 11.0 support
  - Mcirosoft® Windows* HPC Server 2008 & Microsoft® Windows Vista support
  - uDAPL* 2.0 support

*Other names and brands may be claimed as the property of others.

New in Release 3.2 Update 1

- Performance and Scalability enhancements
  - Collective optimizations for newer multicore platforms
  - Scalable mpdboot startup through the -b<maxbranch> and --parallel-startup options
  - (Windows only) Direct interprocess memory copy through the I_MPI_INTRANODE_DIRECT_COPY and I_MPI_INTRANODE_DIRECT_COPY_THRESHOLDS variables

New in release 3.2 Update 1 (cont.)

- Usability improvements
  - (Linux only) Linux* Standard Base (LSB) compliant RPMs
  - ILP64 support through the -ilp64 option
  - Process pinning improvements for newer multicore platforms
  - (Windows only) Active Directory based user authorization through the I_MPI_AUTH_METHOD variable
- Extended interoperability
  - Intel® Compiler 11.1 support
  - SLES11 & RH EL 5.3 support

*Other names and brands may be claimed as the property of others.