
*ISC 2018 05/26/2018*

# Hardware Topologies Working Group

# Working Group Statement

- New architectures increasingly complex
  - Memory/Network hierarchies
- MPI is hardware-agnostic
  - And should remain so
  - But doesn’t prevent nor encourages the application to access the underlying HW (Nodes + Network)
- Issues
  - How to discover HW resources in a MPI application?
  - How to leverage the HW resources?
- Questions
  - What is the right level of abstraction?
  - Which MPI constructs could leverage HW topologies?
  - What are the interactions with other programming models?

# Implicit access to HW topologies

- Current proposal:
  - Creation of so-called hierarchical communicators
    - A communicator corresponds to a specific level in the HW hierarchy
  - Based on the `MPI_Comm_split_type` function
    - Introduce a new split_type value: **`MPI_COMM_TYPE_PHYSICAL_TOPOLOGY`**
- Prototype implementation available: Hsplit
  - External library (for now)
    - Available at : http://mpi-topology.gforge.inria.fr/
  - hwloc/netloc-based implementation

# Explicit access/query of HW topologies

- Determination of processes coordinates/neighborhoods
- MPI_T interface
- Dedicated functions (E.g. Fujitsu's extensions)
  - FJMPI_Topology_get_dimension
  - FJMPI_Topology_rank2x
  - FJMPI_Topology_x2rank
  - ...

# Mapping of parallel applications

- Difficult issue in MPI
  - Currently “Outside the scope of the standard”
  - Involves RJMS, process managers, MPI applications
    - Considering approaches *à la* MPI_Bind
  - Hybrid, dynamic cases
- Not very user-friendly nor portable
- Standardize mpiexec/mpirun parameters?

# Join us!

- We need feedback/uses cases from application developers
  - Mailing list: mpiwg-hw-topology@lists.mpi-forum.org
- Github: https://github.com/mpiwg-hw-topology
  - Teleconferences on regular basis
  - The minutes are available on the WG site