

RMA Working Group Update

Status

- Three proposals submitted over email
- They will be discussed in the working group meeting on Wednesday, 9.30-10.45

Proposal 1

- By Galen Shipman and Howard Pritchard
- Add read-modify-write operations
 - MPI_Rmw for single-operand operations
 - MPI_Rmw2 for two-operand operations

Read-Modify-Write Functions

`MPI_Rmw(void *operand_addr, int count, MPI_Datatype datatype,
void *result_addr, int target_rank, MPI_Aint target_disp,
int assert, MPI_RMW_Op op, MPI_Win win)`

`MPI_Rmw2(void *operand_addr, int count, MPI_Datatype datatype,
void *operand2_addr, void *result_addr, int target_rank,
MPI_Aint target_disp, int assert, MPI_RMW_2_Op op,
MPI_Win win)`

RMW Ops

- MPI_RMW_INC increment
 - MPI_RMW_PROD product
 - MPI_RMW_SUM sum
 - MPI_RMW_LAND logical and
 - MPI_RMW_LOR logical or
 - MPI_RMW_LXOR logical xor
 - MPI_RMW_BAND binary and
 - MPI_RMW_BOR binary or
 - MPI_RMW_BXOR binary xor
 - MPI_RMW_SWAP swap value
-
- MPI_RMW2_MASK_SWAP swap masked bits
 - MPI_RMW2_COMP_SWAP compare and swap

Proposal 2

- By Hubert Ritzdorf
- A new synchronization method to signal the completion of RMA operations at the target
- Similar to the use of “flags” in shared-memory programs to signal completion of data transfer
- Introduces MPI_Sync object
- Persistent functions that return MPI_Request objects
 - MPI_Win_sync_ops_init
 - MPI_Win_sync_object_init
- Can be started with MPI_Start or MPI_Startall and completion can be checked with MPI_Test/Wait

Proposal 3

- By Jesper Larsson Traeff
- Preregister derived datatypes to avoid having to transfer them from origin to target
- `MPI_Type_register(datatype,comm)`
- `MPI_Type_register_sparse(datatype,group,comm)`
- `MPI_Type_dereg(datatype,comm)`
- Same mechanism can be used to extend `MPI_Accumulate` to support user-defined ops
 - `MPI_Op_register(op,comm)`
 - `MPI_Op_register_sparse(op,group,comm)`
 - `MPI_Op_dereg(op,comm)`