

Non-catastrophic Errors

Rank 1

Rank 2

Rank 0

Rank n-1

Allows user to change the # of requests than aborting

MPI_Wait ()

Non-catastrophic Error

of requests > Total requests allowed

1:191 MPI_Isend/MPI_Irecv (OFI/PSM2)

Number of Usable Requests	Rlimit=8	Rlimit=16	Rlimit=32	Rlimit=64	Rlimit=128	Rlimit=256	Rlimit=512	Rlimit=1024	Rlimit=2048
8	100K	100K	100K	100K	100K	100K	100K	100K	100K
16	150K	150K	150K	150K	150K	150K	150K	150K	150K
32	200K	200K	200K	200K	200K	200K	200K	200K	200K
64	250K	250K	250K	250K	250K	250K	250K	250K	250K
128	300K	300K	300K	300K	300K	300K	300K	300K	300K
256	350K	350K	350K	350K	350K	350K	350K	350K	350K
512	400K	400K	400K	400K	400K	400K	400K	400K	400K
1024	450K	450K	450K	450K	450K	450K	450K	450K	450K
2048	450K	450K	450K	450K	450K	450K	450K	450K	450K

Communicator Creation Hints based on Topology Awareness

Communicator 1

Communicator 2

MPI_Comm_split_type (Shared Memory, NUMA)

Communication Behavior Hints

mpi_assert_no_any_source

mpi_assert_allow_overtaking

No overtaking

Allow overtaking

Improved Support for Heterogeneous Memory

Results for the miniFE miniapp on KNL

Runtime

512x256x256 512x512x256 512x512x512

malloc memkind hexe malloc memkind hexe malloc memkind hexe

Mat-struc-gen FE assembly Total CG Other

Optimizations to Virtual Topology Functionality

Overhead (s)

Number of Processes

1K 2K 4K

$\delta=0.05$ $\delta=0.1$ $\delta=0.2$ $\delta=0.4$ $\delta=0.6$ $\delta=0.8$

Phase 1 overheads for the sparse random graph topology across different number of processes

Work-Queue Data Transfer Model

Data Transfer Rate (Chunks/s)

Data Chunk Size (Bytes)

1 4 16 64 256 1024 4096

MS-WorkQ Original