



# PTF: MPI-Parameters Plugin

Eduardo César, Gertvjola Saveta

[eduardo.cesar@uab.cat](mailto:eduardo.cesar@uab.cat)

[gertvjola.saveta@caos.uab.es](mailto:gertvjola.saveta@caos.uab.es)

February 13th 2014

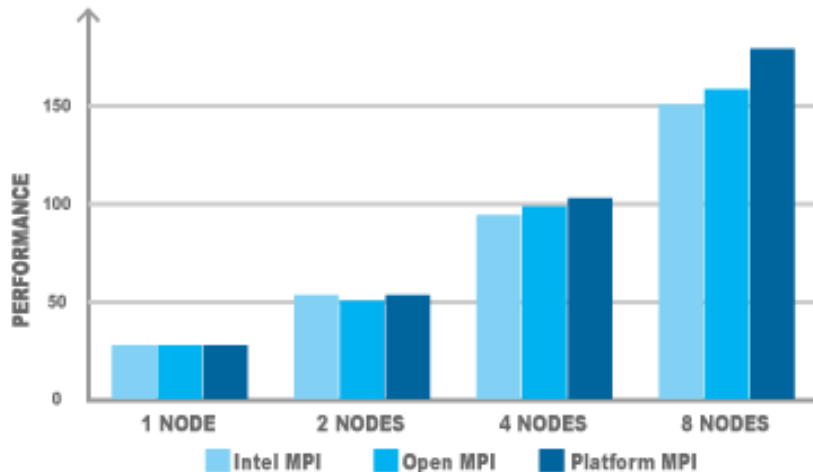


Technische Universität München

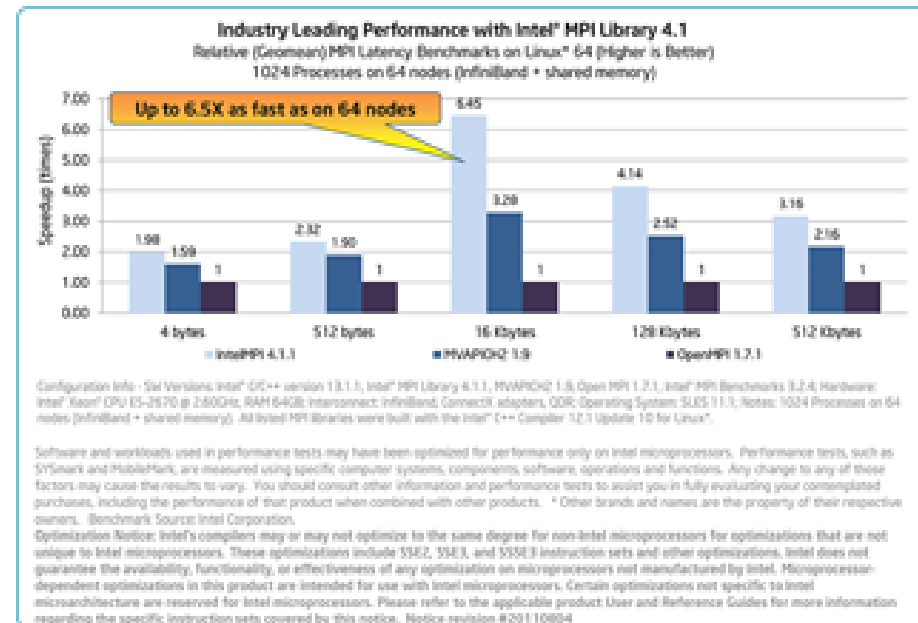


- **IBM MPI, Intel MPI, OpenMPI**: from more than 50 to more than 150 parameters, **from ten to several tens of them can influence performance**
- Manual tuning can be very **difficult** and extremely **time consuming**
- Tuning library parameters is so important that Intel MPI and OpenMPI include their own tuning tools (mpitune and OTPO)

All vendors claim they offer the best performance!!



Source: MILC Benchmark Test, HPC Advisory Council, Sept 2012



Performance can only be obtained from a good tuning of library parameters!!

- **Processor affinity/ Process Pinning**

Binding processes to certain processors can improve performance by inhibiting excessive process movement, reducing resource contention, or improving interprocess communications.

It can also degrade performance by inhibiting the OS capability to balance loads.

Ex.

MCA\_orte\_process\_binding (OpenMPI)

I\_MPI\_PROCESSOR\_LIST (IntelMPI)

MP\_TASK\_AFFINITY (IBM MPI)

- **Memory affinity** (NUMA architectures)

Allocating all memory local to the processor that asked for it. Improving performance when shared memory is used for communication

- **Message control & Buffer size**

Determining the values of send/recv buffers and message sizes for using different communication protocols on different fabrics

Determining collective communication algorithms

Ex.

MP\_EAGER\_LIMIT (IBM MPI)

I\_MPI\_EAGER\_THRESHOLD (Intel MPI)

MCA\_btl\_openib\_eager\_limit (OpenMPI)

- **Functionality:**

Exhaustive search using all the environment variables and list/ranges of values specified by the user.

The combination of values leading to the lowest execution time is given to the user as an advise

- **Configuration**

Variable/parameter names and possible values are specified in a configuration file (`PSC_PARAM_SPEC_FILE` || `./param_spec.conf`)

MPI implementation is also indicated in the configuration file (supporting IBM MPI and OpenMPI)

- Configuration File Format:

```
<configuration-file-start *+ [Library Implementation]>
```

```
MPIPO_BEGIN [MPI_OPENMPI || MPI_IBM]
```

```
<parameter-name=initial-value : [+|*]step : final-value>; |
```

```
<parameter-name=<comma-separated-list-of-values>;
```

```
....
```

```
<configuration-file-end>
```

```
ENDLS
```



- Example

```
MPIPO_BEGIN openmpi
```

```
mp_buffer_mem=16M,32M,64M;
```

```
mp_eager_limit=12:2:16;
```

```
MPIPO_END
```

## Load required models :

- › **module use ~nct00001/gpfs\_projects/UNITE/tutorial/mf**
- › **module load UNITE**  
UNITE loaded
- › **module load periscope**  
load PYTHON/2.7.3 (PATH, MANPATH, LD\_LIBRARY\_PATH,C\_INCLUDE\_PATH)  
load BOOST/1.52.0 (PATH, LD\_LIBRARY\_PATH, NETCDF)

## Copy to your home:

- › `cd ~`
- › `cp -r ~nct00001/gpfs_projects/thursday_material/mpiparameters .`
- › `mv mpiparameters/periscope.sample .periscope`
- › `cd mpiparameters`
- › `unzip Mpiparameters_Demo.zip`

The configuration file is pre-difined in :

- cd NPB3.3-MZ-MPI
- make suite
- cd bin
- ls  
param\_spec.conf

```
MPIPO_BEGIN openmpi
```

```
mp_buffer_mem=16M,32M,64M;  
mp_eager_limit=12:2:16;
```

```
MPIPO_END
```

### Command line for the mpi parameters tuning

- `psc_frontend --apprun=./bt-mz.W.4 --uninstrumented --mpinumprocs=4 --tune=mpiparameters`
- Use the job script
- `bsub < job.lsf`

## Found best scenario: 8

Parameter combination:

`--mca mp_buffer_mem 64M --mca mp_eager_limit 16`

All Results:

Scenario	Runtime	Flags
0	2.359510	--mca mp_buffer_mem 16M --mca mp_eager_limit 12
1	2.251537	--mca mp_buffer_mem 16M --mca mp_eager_limit 14
2	2.243806	--mca mp_buffer_mem 16M --mca mp_eager_limit 16
3	2.224361	--mca mp_buffer_mem 32M --mca mp_eager_limit 12
4	2.240855	--mca mp_buffer_mem 32M --mca mp_eager_limit 14
5	2.226031	--mca mp_buffer_mem 32M --mca mp_eager_limit 16
6	2.221029	--mca mp_buffer_mem 64M --mca mp_eager_limit 12
7	2.230055	--mca mp_buffer_mem 64M --mca mp_eager_limit 14
8	2.218171	--mca mp_buffer_mem 64M --mca mp_eager_limit 16

-----

End Periscope run! Search took 33.2387 seconds ( 0 seconds for startup )

-----

[psc\_frontend][INFO:fe] Experiment completed!

[psc\_frontend][INFO:fe] Exporting results to properties\_tune\_4484.psc

Any Questions?