Hands-on: *Archer* Cray XC30
NPB-MZ-MPI / bt-mz_C.8

VI-HPS Team

Tutorial exercise objectives

- Familiarise with usage of VI-HPS tools
  - complementary tools' capabilities & interoperability
- Prepare to apply tools productively to *your* applications(s)
- Exercise is based on a small portable benchmark code
  - unlikely to have significant optimisation opportunities

- Optional (recommended) exercise extensions
  - analyse performance of alternative configurations
  - investigate effectiveness of system-specific compiler/MPI optimisations and/or placement/binding/affinity capabilities
  - investigate scalability and analyse scalability limiters
  - compare performance on different HPC platforms
  - …

Setup for exercises (Archer Cray XC30)

- Connect to your (ta002 training) account on Archer (with X11-forwarding enabled)

```
% ssh –X login.archer.ac.uk –l <USER>
```

- Select appropriate PrgEnv (cray, gnu, or intel: gnu recommended/assumed)

```
% module switch PrgEnv-cray PrgEnv-gnu
```

- Copy tutorial sources to your WORK directory

```
% cd /work/ta002/ta002/$USER
% tar zxvf  /work/ta002/shared/tutorial/NPB3.3-MZ-MPI.tar.gz
% cd NPB3.3-MZ-MPI
```

NPB-MZ-MPI Suite

- The NAS Parallel Benchmark suite (MPI+OpenMP version)
  - Available from:

    ## http://www.nas.nasa.gov/Software/NPB

  - 3 benchmarks in Fortran77
  - Configurable for various sizes & classes
- Move into the NPB3.3-MZ-MPI root directory

```
% ls
bin/      common/   jobscript/   Makefile   README.install   SP-MZ/
BT-MZ/  config/   LU-MZ/        README     README.tutorial  sys/
```

- Subdirectories contain source code for each benchmark
  - plus additional configuration and common code
- The provided distribution has already been configured for the tutorial, such that it is ready to "make" one or more of the benchmarks
  - but config/make.def may first need to be adjusted to specify appropriate PrgEnv compiler flags

NPB-MZ-MPI / BT: config/make.def

```
#              SITE- AND/OR PLATFORM-SPECIFIC DEFINITIONS.
#
#---------------------------------------------------------------------

#---------------------------------------------------------------------
# Configured for Cray with PrgEnv compiler-specific OpenMP flags
#---------------------------------------------------------------------
#OPENMP = -homp            # Cray/CCE compiler
OPENMP = -fopenmp         # GNU/GCC compiler
#OPENMP = -qopenmp         # Intel compiler

...
#---------------------------------------------------------------------
# The Fortran compiler used for MPI programs
#---------------------------------------------------------------------
MPIF77 = ftn

# Alternative variant to perform instrumentation
#MPIF77 = scorep --user  ftn

# PREP is a generic preposition macro for instrumentation preparation
#MPIF77 = $(PREP) ftn
...
```

Uncomment COMPILER flags according to current PrgEnv

Default (no instrumentation)

Hint: uncomment a compiler wrapper to do instrumentation

Building an NPB-MZ-MPI Benchmark

```
% make
   ===========================================
   =      NAS PARALLEL BENCHMARKS 3.3      =
   =      MPI+OpenMP Multi-Zone Versions    =
   =      F77                               =
   ===========================================


   To make a NAS multi-zone benchmark type

           make <benchmark-name> CLASS=<class> NPROCS=<nprocs>

   where <benchmark-name> is "bt-mz", "lu-mz", or "sp-mz"
         <class>            is "S", "W", "A" through "F"
         <nprocs>           is number of processes


    [...]


   ******************************************************
   * Custom build configuration is specified in config/make.def  *
   * Suggested tutorial exercise configuration for Archer:       *
   *        make bt-mz CLASS=C NPROCS=8                           *
   ******************************************************
```

- Type "make" for instructions

Building an NPB-MZ-MPI Benchmark

```
% make bt-mz CLASS=C NPROCS=8
make[1]: Entering directory `BT-MZ'
make[2]: Entering directory `sys'
cc  -o setparams setparams.c -lm
make[2]: Leaving directory `sys'
../sys/setparams bt-mz 8 C
make[2]: Entering directory `../BT-MZ'
ftn  -c  -O3 -fopenmp       bt.f
[…]
ftn  -c  -O3 -fopenmp       mpi_setup.f
cd ../common;  ftn  -c  -O3 -fopenmp       print_results.f
cd ../common;  ftn  -c  -O3 -fopenmp       timers.f
ftn  -O3 -fopenmp   -o ../bin/bt-mz_C.8 bt.o
 initialize.o exact_solution.o exact_rhs.o set_constants.o adi.o
 rhs.o zone_setup.o x_solve.o y_solve.o  exch_qbc.o solve_subs.o
 z_solve.o add.o error.o verify.o mpi_setup.o ../common/print_results.o
 ../common/timers.o
make[2]: Leaving directory `BT-MZ'
Built executable ../bin/bt-mz_C.8
make[1]: Leaving directory `BT-MZ'
```

- Specify the benchmark configuration
  - benchmark name: **bt-mz**, lu-mz, sp-mz
  - the benchmark class (S, W, A, B, C, D, E): CLASS=**C**
  - the number of MPI processes: NPROCS=**8**

Shortcut: `% make suite`

NPB-MZ-MPI / BT (Block Tridiagonal Solver)

- What does it do?
  - Solves a discretized version of the unsteady, compressible Navier-Stokes equations in three spatial dimensions
  - Performs 200 time-steps on a regular 3-dimensional grid
- Implemented in 20 or so Fortran77 source modules

- Uses MPI & OpenMP in combination
  - 8 processes each with 6 threads should be reasonable for 2 compute nodes of Archer
  - bt-mz_C.8 should run in
    - around 20 seconds with PrgEnv-gnu
    - around 17 seconds with PrgEnv-cray
    - around 77 seconds with PrgEnv-intel

NPB-MZ-MPI / BT Reference Execution

```
%  cd bin
%  cp ../jobscript/archer/run.pbs .
%  less run.pbs
%  qsub run.pbs

%  cat mzmpibt.o<job_id>
 NAS Parallel Benchmarks (NPB3.3-MZ-MPI) – BT-MZ MPI+OpenMP Benchmark

 Number of zones:  16 x  16
 Iterations: 200   dt:   0.000100
 Number of active processes:     8
 Total number of threads:     48  (  6.0 threads/process)

 Time step    1
 Time step   20
  [...]
 Time step  180
 Time step  200
 Verification Successful

 BT-MZ Benchmark Completed.
 Time in seconds = 19.57
```

- Copy jobscript and launch as a hybrid MPI+OpenMP application

Hint: save the benchmark output (or note the run time) to be able to refer to it later