Profiling and Tracing HPC Applications on Comet

Presented by
Mahidhar Tatineni, SDSC
Nov 2, 2018

MPI – Profiling, Tracing Tools

- Several options available. On Comet we have mpiP, TAU, and Allinea MAP installed.
- Useful when you are trying to isolate performance issues.
- Tools can give you info on how much time is being spent in communication. The levels of detail vary with each tool.
- In general identify scaling bottlenecks and try to overlap communication with computation where possible.

mpiP

- **Source: http://mpip.sourceforge.net/**
- Lightweight profiling library for MPI applications
- Low overhead, task local for the most part
- MPI time, call site, and message size information.
- Good tool to get a quick summary of your code.

mpiP example

- **Location: /home/$USER/PROFILING/NPB3.3-MPI**
- **Instructions in the file *mpiP.commands.txt***
- **Compile:**

```
cd /home/$USER/PROFILING/NPB3.3-MPI
cp config/make.def.mpiP config/make.def
make clean
make SP NPROCS=64 CLASS=C
```

- **Run:**

```
cd /home/$USER/PROFILING/NPB3.3-MPI/bin
sbatch --res=UCRES SP-mpip.sb
```

- **Once the job runs you get a .mpiP file.**

Compilation with mpiP

```
#---------------------------------------------------------------------------
# These macros are passed to the linker to help link with MPI correctly
#---------------------------------------------------------------------------
FMPI_LIB  = -L/share/apps/compute/mpiP/v3.4.1/mv2/lib -lmpiP -L/share/apps/compu
te/libiberty -liberty -L/share/apps/compute/libunwind/v1.1/mv2/lib -lunwind
```

```
mpif77 -c  -O3  ninvr.f
mpif77 -c  -O3  y_solve.f
mpif77 -c  -O3  pinvr.f
mpif77 -c  -O3  z_solve.f
mpif77 -c  -O3  tzetar.f
mpif77 -c  -O3  add.f
mpif77 -c  -O3  txinvr.f
mpif77 -c  -O3  error.f
mpif77 -c  -O3  verify.f
mpif77 -c  -O3  setup_mpi.f
cd ../common; mpif77 -c  -O3  print_results.f
cd ../common; mpif77 -c  -O3  timers.f
mpif77 -O3  -o ../bin/sp.C.64 sp.o make_set.o initialize.o exact_solution.o exac
t_rhs.o set_constants.o adi.o define.o copy_faces.o rhs.o lhsx.o lhsy.o lhsz.o x
_solve.o ninvr.o y_solve.o pinvr.o z_solve.o tzetar.o add.o txinvr.o error.o ver
ify.o setup_mpi.o ../common/print_results.o ../common/timers.o -L/share/apps/com
pute/mpiP/v3.4.1/mv2/lib -lmpiP -L/share/apps/compute/libiberty -liberty -L/shar
e/apps/compute/libunwind/v1.1/mv2/lib -lunwind
make[1]: Leaving directory `/home/etrain63/PROFILING/NPB3.3-MPI/SP'
```

mpiP output (in std out)

```
[etrain63@comet-ln2 bin]$ more SP-mpip.19924535.comet-13-16.out
mpiP:
mpiP: mpiP: mpiP V3.4.1 (Build Aug  9 2018/12:07:41)
mpiP: Direct questions and errors to mpip-help@lists.sourceforge.net
mpiP:


 NAS Parallel Benchmarks 3.3 -- SP Benchmark

 No input file inputsp.data. Using compiled defaults
 Size:  162x 162x 162
 Iterations:  400    dt:   0.0006700
 Number of active processes:    64
```

```
 Please send feedbacks and/or the results of this run to:

 NPB Development Team
 Internet: npb@nas.nasa.gov


mpiP:
mpiP: Storing mpiP output in [./sp.C.64.64.24561.1.mpiP].
mpiP:
```

mpiP Output
Date, Time + Task Assignment

```
@ mpiP
@ Command : /home/etrain63/PROFILING/NPB3.3-MPI/bin/sp.C.64
@ Version                  : 3.4.1
@ MPIP Build date          : Aug  9 2018, 12:07:41
@ Start time               : 2018 11 01 17:55:26
@ Stop time                : 2018 11 01 17:55:43
@ Timer Used               : PMPI_Wtime
@ MPIP env var             : [null]
@ Collector Rank           : 0
@ Collector PID            : 24561
@ Final Output Dir         : .
@ Report generation        : Single collector task
@ MPI Task Assignment      : 0 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 1 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 2 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 3 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 4 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 5 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 6 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 7 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 8 comet-13-16.sdsc.edu
@ MPI Task Assignment      : 9 comet-13-16.sdsc.edu
```

mpiP output
Total, MPI Time per task

```
---------------------------------------------------------------
@--- MPI Time (seconds) ----------------------------------------
---------------------------------------------------------------
Task    AppTime    MPITime     MPI%
   0       17.7       2.51    14.19
   1       17.7       2.49    14.06
   2       17.7       2.59    14.65
   3       17.7       2.68    15.14
   4       17.7       2.38    13.42
   5       17.7       2.66    15.00
   6       17.7       2.39    13.49
   7       17.7        2.4    13.55
   8       17.7       2.74    15.51
   9       17.7        3.1    17.51
  10       17.7       2.86    16.17
  11       17.7       2.88    16.29
  12       17.7       2.76    15.60
  13       17.7       3.13    17.71
  14       17.7       2.69    15.21
  15       17.7       2.98    16.82
  16       17.7       2.42    13.68
  17       17.7       2.66    15.04
  18       17.7       2.65    14.95
  19       17.7       2.75    15.54
```

```
  55       17.6       2.17    12.33
  56       17.6       3.51    19.90
  57       17.6       2.43    13.80
  58       17.6       3.55    20.11
  59       17.6       2.02    11.43
  60       17.6       3.36    19.08
  61       17.6       2.08    11.80
  62       17.6       3.46    19.62
  63       17.6       2.21    12.55
   *   1.13e+03        180    15.89
```

mpiP output: MPI Call Summary

```
---------------------------------------------------------------------------
@--- Aggregate Time (top twenty, descending, milliseconds) ----------------
---------------------------------------------------------------------------
Call                 Site       Time    App%    MPI%      COV
Waitall                 2   4.75e+04    4.20   26.43    0.42
Waitall                 4   3.84e+04    3.39   21.37    0.27
Waitall                 1   3.24e+04    2.86   18.03    0.29
Waitall                 7   3.07e+04    2.72   17.12    0.36
Isend                   7   6.23e+03    0.55    3.47    0.52
Isend                   2   5.06e+03    0.45    2.82    0.08
Isend                   1   4.93e+03    0.44    2.75    0.15
Isend                   4   4.17e+03    0.37    2.32    0.08
Irecv                   4   2.31e+03    0.20    1.29    0.10
Irecv                   1   2.29e+03    0.20    1.28    0.11
Irecv                   2   2.22e+03    0.20    1.23    0.08
Comm_split              9   2.01e+03    0.18    1.12    0.91
Irecv                   7   1.03e+03    0.09    0.58    0.07
Allreduce               5       98.6    0.01    0.05    0.57
Waitall                 8       79.7    0.01    0.04    0.36
Barrier                 6       52.7    0.00    0.03    0.41
Allreduce               3       49.8    0.00    0.03    0.45
Bcast                   6       30.3    0.00    0.02    0.11
Isend                   8       18.6    0.00    0.01    0.66
Comm_dup                9       12.8    0.00    0.01    0.18
```

mpiP output
Callsite Details

```
---------------------------------------------------------------------------
@--- Callsite Time statistics (all, milliseconds): 1408 --------------------
---------------------------------------------------------------------------
Name                 Site Rank  Count      Max     Mean      Min   App%   MPI%
Allreduce               3    0      1     0.67     0.67     0.67   0.00   0.03
Allreduce               3    1      1     0.31     0.31     0.31   0.00   0.01
Allreduce               3    2      1    0.088    0.088    0.088   0.00   0.00
Allreduce               3    3      1    0.941    0.941    0.941   0.01   0.04
Allreduce               3    4      1    0.497    0.497    0.497   0.00   0.02
Allreduce               3    5      1    0.986    0.986    0.986   0.01   0.04
Allreduce               3    6      1    0.953    0.953    0.953   0.01   0.04
Allreduce               3    7      1    0.588    0.588    0.588   0.00   0.02
Allreduce               3    8      1    0.353    0.353    0.353   0.00   0.01
Allreduce               3    9      1    0.218    0.218    0.218   0.00   0.01
Allreduce               3   10      1    0.812    0.812    0.812   0.00   0.03
Allreduce               3   11      1    0.559    0.559    0.559   0.00   0.02
Allreduce               3   12      1     0.97     0.97     0.97   0.01   0.04
Allreduce               3   13      1     1.02     1.02     1.02   0.01   0.03
Allreduce               3   14      1    0.623    0.623    0.623   0.00   0.02
Allreduce               3   15      1     1.01     1.01     1.01   0.01   0.03
```

Instrumentation using TAU

- Dynamic with library preloading
  - No recompilation / source code modification required
  - Uses tau_exec utility
  - Flags determine the scope of instrumentation

- Compiler based
  - Need to recompile with TAU wrappers
  - Uses tau_f90.sh, tau_cc.sh, and tau_cxx.sh wrappers

- We will see both examples today

Compilation using mpicc, mpif77
(For TAU Instrumentation using Interposition)

- **Change to the directory:**
  - cd /home/$USER/PROFILING/NPB3.3-MPI
- **Copy the config file:**
  - cp config/make.def.mpicc config/make.def
- **Compile:**
  - make SP NPROCS=64 CLASS=C
  - SP: Benchmark name
  - NPROCS: Number of MPI tasks we will be using
  - CLASS: Size of the benchmark

TAU: Instrumentation using Interposition

- Dynamic instrumentation is via library pre-loading.
- Run via ***tau_exec***
- Command-line options to ***tau_exec***

```
#!/bin/bash
#SBATCH --job-name="example1"
#SBATCH --output="example1.%j.%N.out"
#SBATCH --partition=compute
#SBATCH --nodes=3
#SBATCH --ntasks-per-node=24
#SBATCH --export=ALL
#SBATCH -t 01:30:00
module load tau
export TAU_METRICS=PAPI_LST_INS:PAPI_L1_DCM
export SLURM_NODEFILE=`generate_pbs_nodefile`
mpirun_rsh -export-all -hostfile $SLURM_NODEFILE -np 64 tau_exec -T
papi,mpi,pdt ./sp.C.64
```

TAU: Instrumentation using Interposition

- **Submit the job:**
  - cd bin
  - sbatch --res=UCRES tau-example1.sb
- **Check the job:**
  - squeue -u $USER
- **Once the run is complete, check output:**

```
[etrain63@comet-ln2 bin]$ ls -lt
total 48
-rw-r--r--  1 etrain63 gue998   2331 Nov  1 21:21 example1.19926762.comet-02-61.out
drwxr-x---  2 etrain63 gue998     66 Nov  1 21:21 MULTI__PAPI_L1_DCM
drwxr-x---  2 etrain63 gue998     66 Nov  1 21:21 MULTI__PAPI_LST_INS
```

Compilation using tau_cc.sh, tau_f77.sh

- **Change the config file:**

  cp config/make.def.tau config/make.def

- **Set the TAU Makefile:**

  export TAU_MAKEFILE=$TAUHOME/x86_64/lib/Makefile.tau-icpc-papi-mpi-pdt

- **Clean the previous build and recompile:**

  make clean

  make SP NPROCS=64 CLASS=C

- **Submit the Job:**

  cd bin

  sbatch --res=UCRES tau-example2.sb

Example 2: Output

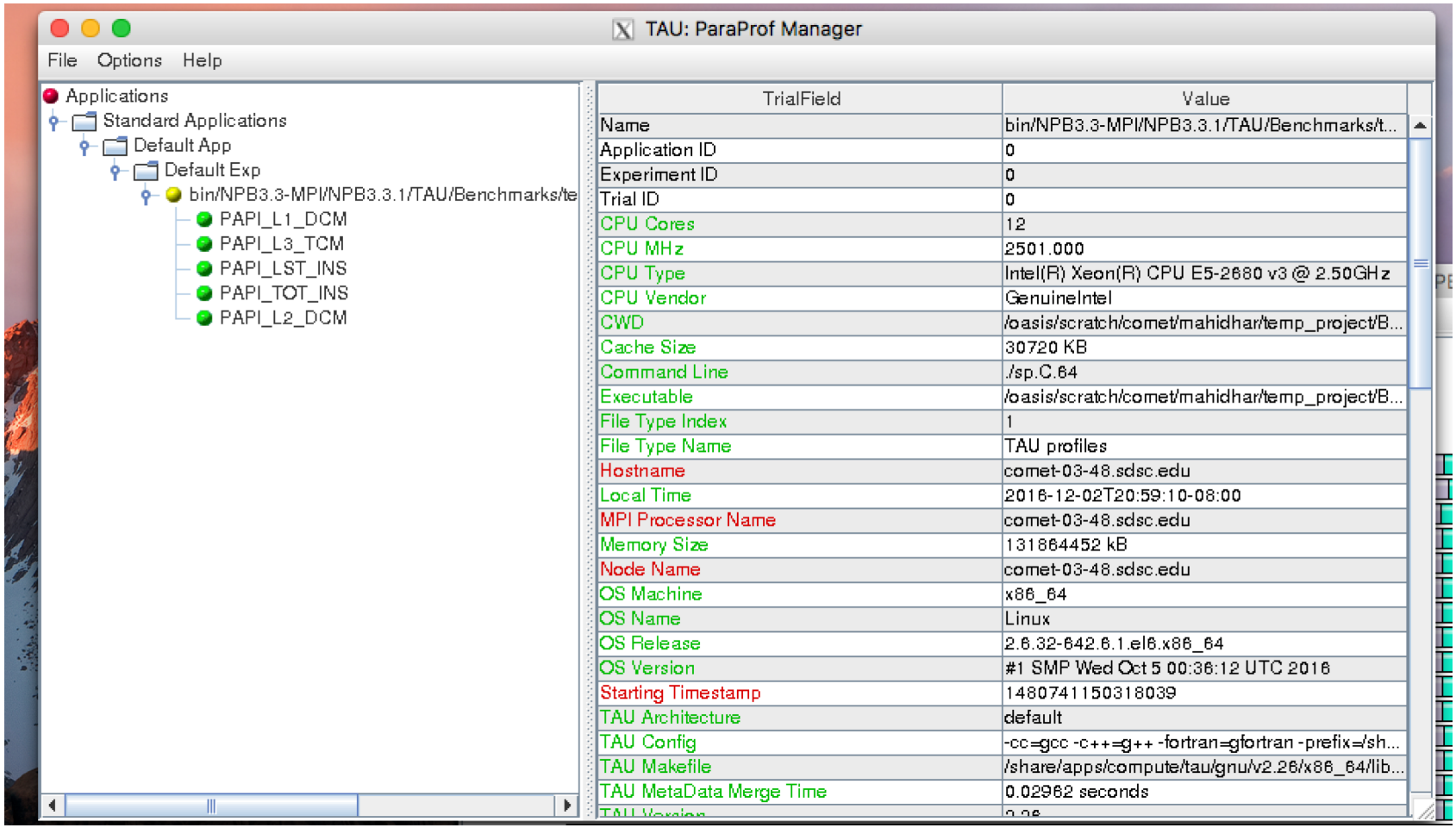
```
[etrain63@comet-ln2 bin]$ ls -lt
total 3494
    drwxr-x--- 2 etrain63 gue998   66 Nov  1 21:02 MULTI__PAPI_TOT_INS
    -rw-r--r-- 1 etrain63 gue998 2367 Nov  1 21:02 example2.19926404.comet-23-08.out
    drwxr-x--- 2 etrain63 gue998   66 Nov  1 21:02 MULTI__PAPI_LST_INS
    drwxr-x--- 2 etrain63 gue998   66 Nov  1 21:02 MULTI__PAPI_L3_TCM
    drwxr-x--- 2 etrain63 gue998   66 Nov  1 21:02 MULTI__PAPI_L2_DCM
    drwxr-x--- 2 etrain63 gue998   66 Nov  1 21:02 MULTI__PAPI_L1_DCM
```

- If X is enabled we can use ParaProf to study the output
- Can be very slow given the latencies - snapshots presented in next few slides.

pprof – text based analysis

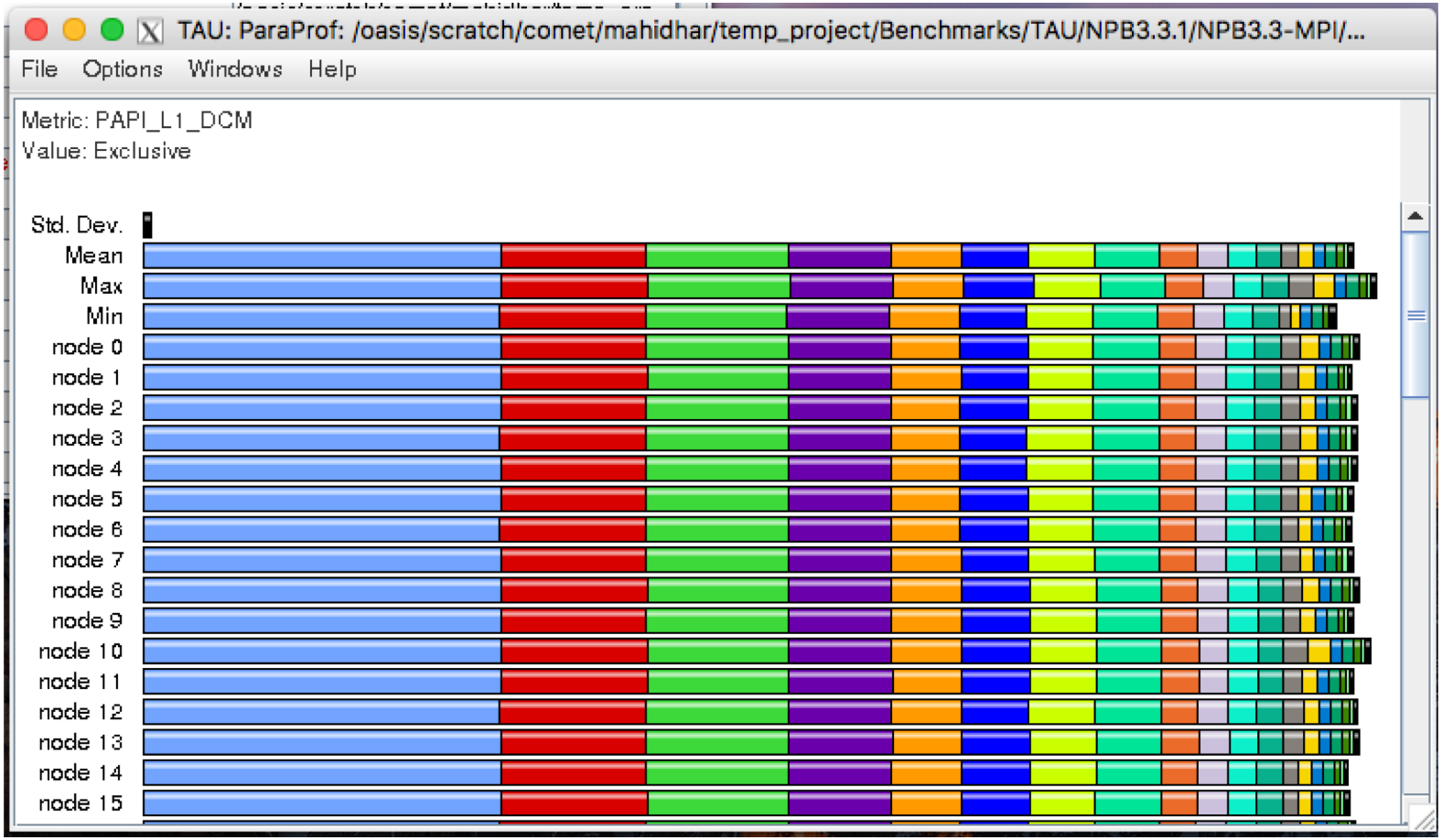
```
[etrain63@comet-ln2 MULTI__PAPI_TOT_INS]$ pprof | more
Reading Profile files in profile.*

NODE 0;CONTEXT 0;THREAD 0:
---------------------------------------------------------------------------------------
%Time    Exclusive    Inclusive       #Call      #Subrs Count/Call Name
              counts total counts
---------------------------------------------------------------------------------------
100.0    6.04E+05    6.44E+10           1         443 64395147131 MPSP
 87.7   2.652E+06   5.646E+10         401        2406  140790410 ADI
 29.9   1.279E+09   1.924E+10         402        5628   47871162 COPY_FACES
 27.7   6.404E+09   1.785E+10         401       23258   44511774 X_SOLVE
 24.6   1.584E+10   1.584E+10         402           0   39405908 COMPUTE_RHS
 22.3   1.436E+10   1.436E+10       17244           0     832939 MPI_Waitall()
 14.3   4.616E+09   9.193E+09         401       23258   22926058 Z_SOLVE
 13.1   4.466E+09   8.409E+09         401       23258   20969563 Y_SOLVE
  6.9   4.467E+09   4.467E+09           2           0 2233532306 MPI_Barrier()
  2.9   1.865E+09   1.865E+09        3208           0     581278 LHSX
  2.9   1.846E+09   1.846E+09        3208           0     575528 LHSZ
  2.9   1.842E+09   1.842E+09        3208           0     574303 LHSY
  2.3   8.072E+08   1.506E+09           2      802980  752898516 INITIALIZE
  1.7   1.123E+09   1.123E+09        3208           0     350169 TZETAR
  1.7   1.099E+09   1.099E+09         401           0    2740206 TXINVR
  1.5   9.546E+08   9.546E+08 1.09729E+06           0        870 EXACT_SOLUTION
  1.4   9.615E+04   8.776E+08           1           3  877622067 VERIFY
  1.3   6.671E+07   8.213E+08           1       66482  821313917 ERROR_NORM
  1.1     7.1E+08     7.1E+08         401           0    1770485 ADD
  1.1   7.016E+08   7.016E+08           2           0  350824942 MPI_Allreduce()
  0.9   3.201E+04   5.677E+08           1           7  567747978 SETUP_MPI
  0.7    2.72E+08   4.702E+08           1      227833  470244405 EXACT_RHS
  0.7   4.439E+08   4.439E+08        3208           0     138364 PINVR
  0.7   4.439E+08   4.439E+08        3208           0     138364 NINVR
  0.7   4.326E+08   4.326E+08           1           0  432629333 MPI_Init()
  0.2   1.342E+08   1.342E+08           1           0  134201649 MPI_Comm_split()
  0.2   1.122E+08   1.122E+08       19254           0       5827 MPI_Isend()
  0.1   5.015E+07   5.015E+07       19254           0       2605 MPI_Irecv()
  0.1   4.476E+07   4.476E+07           1           0   44758662 MPI_Finalize()
  0.0   1.271E+06   6.155E+06           1           1    6155206 RHS_NORM
  0.0   4.018E+06   4.018E+06           1           0    4017997 LHSINIT
  0.0   8.826E+05   8.826E+05           2           0     441292 MPI_Comm_dup()
  0.0   1.456E+05   1.456E+05           1           0     145647 MPI_Reduce()
  0.0   1.169E+05   1.169E+05           4           0      29223 MPI_Bcast()
  0.0   4.421E+04   4.421E+04           1           0      44212 PRINT_RESULTS
  0.0   1.577E+04   1.577E+04          21           0        751 TIMER_CLEAR
  0.0        7773        7773           1           0       7773 TIMER_START
  0.0        1624        1624           1           0       1624 MAKE_SET
  0.0        1597        1597           2           0        798 MPI_Comm_size()
  0.0        1203        1203           1           0       1203 COMPUTE_BUFFER_SIZE
  0.0        1058        1058           1           0       1058 SET_CONSTANTS
  0.0         875         875           1           0        875 TIMER_STOP
  0.0         804         804           1           0        804 MPI_Comm_rank()
  0.0         751         751           1           0        751 TIMER_READ
---------------------------------------------------------------------------------------

USER EVENTS Profile :NODE 0, CONTEXT 0, THREAD 0
---------------------------------------------------------------------------------------
NumSamples   MaxValue   MinValue  MeanValue  Std. Dev.  Event Name
---------------------------------------------------------------------------------------
         2         40         40         40          0  Message size for all-reduce
         4         12          4          7      3.317  Message size for broadcast
         1          8          8          8          0  Message size for reduce
         2         12          8         10          2  [GROUP=MAX_MARKER] Message size for broadcast
         2         12          8         10          2  [GROUP=MAX_MARKER] Message size for broadcast : MPSP => MPI_Bcast()
---------------------------------------------------------------------------------------
```

ParaProf: NAS SP Benchmark

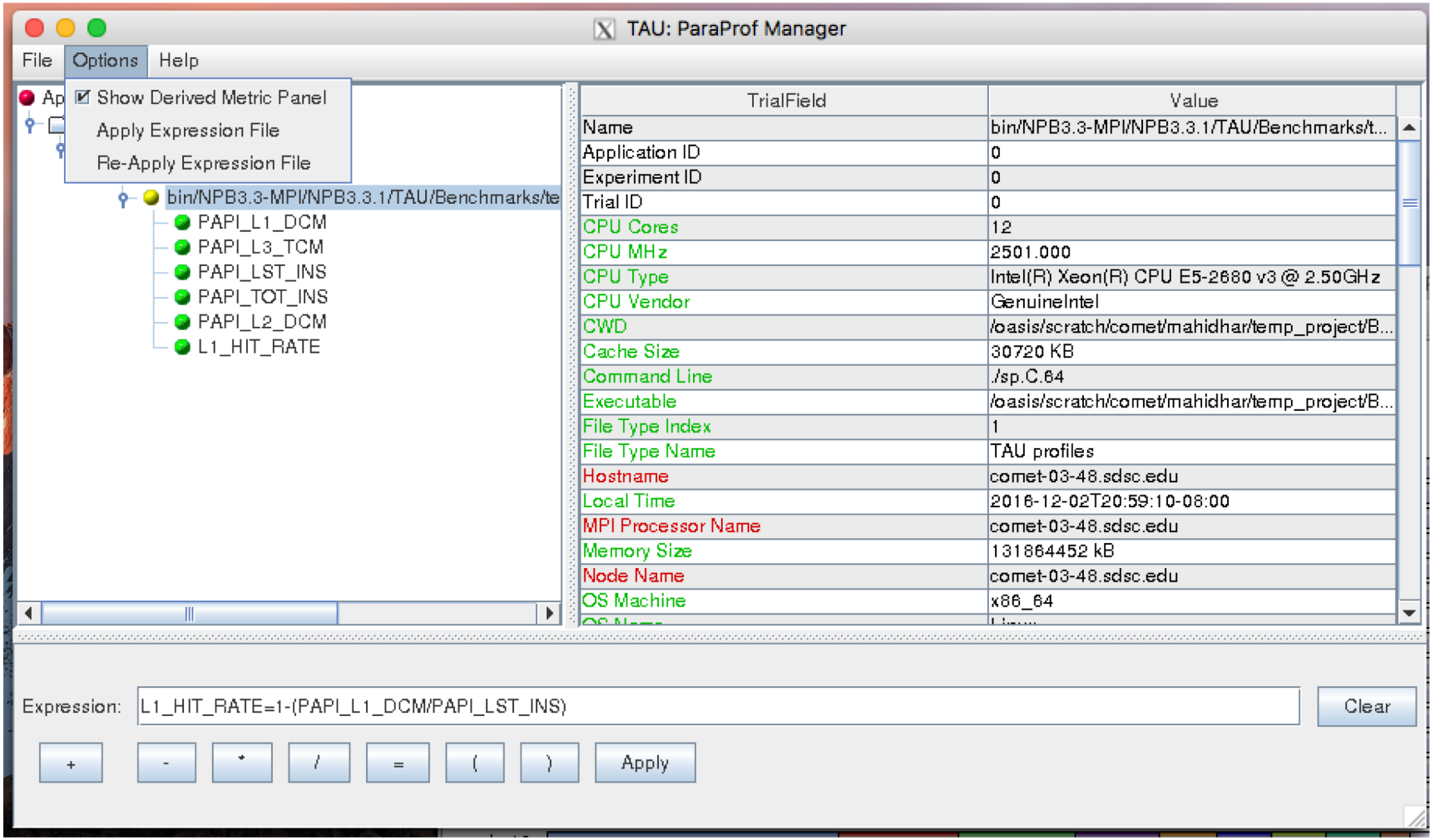


TrialField	Value
Name	bin/NPB3.3-MPI/NPB3.3.1/TAU/Benchmarks/t...
Application ID	0
Experiment ID	0
Trial ID	0
CPU Cores	12
CPU MHz	2501.000
CPU Type	Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz
CPU Vendor	GenuineIntel
CWD	/oasis/scratch/comet/mahidhar/temp_project/B...
Cache Size	30720 KB
Command Line	./sp.C.64
Executable	/oasis/scratch/comet/mahidhar/temp_project/B...
File Type Index	1
File Type Name	TAU profiles
Hostname	comet-03-48.sdsc.edu
Local Time	2016-12-02T20:59:10-08:00
MPI Processor Name	comet-03-48.sdsc.edu
Memory Size	131864452 kB
Node Name	comet-03-48.sdsc.edu
OS Machine	x86_64
OS Name	Linux
OS Release	2.6.32-642.6.1.el6.x86_64
OS Version	#1 SMP Wed Oct 5 00:36:12 UTC 2016
Starting Timestamp	1480741150318039
TAU Architecture	default
TAU Config	-cc=gcc -c++=g++ -fortran=gfortran -prefix=/sh...
TAU Makefile	/share/apps/compute/tau/gnu/v2.26/x86_64/lib...
TAU MetaData Merge Time	0.02962 seconds

PAPI_L1_DCM Metric
NAS SP Benchmark, 64 cores, Class C

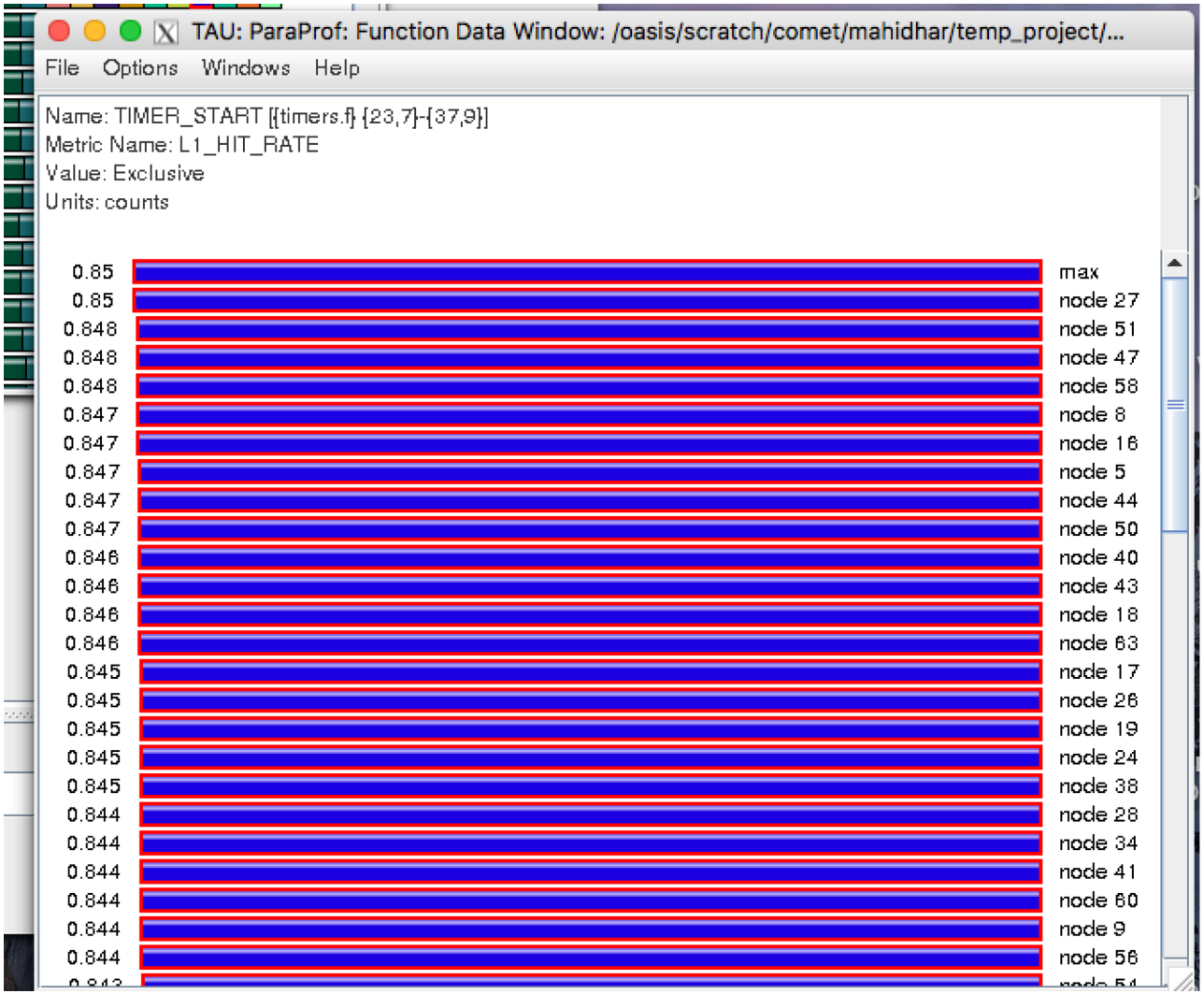


TAU: Derived Metrics



TAU: Derived Metric, L1_HIT_RATE

TAU: Derived Metric L1_HIT_RATE



Utility: papi_avail

```
[mahidhar@comet-ln2 ~]$ papi_avail
Available PAPI preset and user defined events plus hardware information.
--------------------------------------------------------------------------------
PAPI Version             : 5.5.1.0
Vendor string and code   : GenuineIntel (1)
Model string and code    : Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz (63)
CPU Revision             : 2.000000
CPUID Info               : Family: 6  Model: 63  Stepping: 2
CPU Max Megahertz        : 2501
CPU Min Megahertz        : 1200
Hdw Threads per core     : 1
Cores per Socket         : 12
Sockets                  : 2
NUMA Nodes               : 2
CPUs per Node            : 12
Total CPUs               : 24
Running in a VM          : no
Number Hardware Counters : 11
Max Multiplex Counters   : 32
--------------------------------------------------------------------------------


================================================================================
  PAPI Preset Events
================================================================================
    Name        Code    Avail Deriv Description (Note)
PAPI_L1_DCM  0x80000000  Yes   No   Level 1 data cache misses
PAPI_L1_ICM  0x80000001  Yes   No   Level 1 instruction cache misses
PAPI_L2_DCM  0x80000002  Yes   Yes  Level 2 data cache misses
PAPI_L2_ICM  0x80000003  Yes   No   Level 2 instruction cache misses
PAPI_L3_DCM  0x80000004  No    No   Level 3 data cache misses
PAPI_L3_ICM  0x80000005  No    No   Level 3 instruction cache misses
PAPI_L1_TCM  0x80000006  Yes   Yes  Level 1 cache misses
PAPI_L2_TCM  0x80000007  Yes   No   Level 2 cache misses
PAPI_L3_TCM  0x80000008  Yes   No   Level 3 cache misses
PAPI_CA_SNP  0x80000009  Yes   No   Requests for a snoop
PAPI_CA_SHR  0x8000000a  Yes   No   Requests for exclusive access to shared cache line
PAPI_CA_CLN  0x8000000b  Yes   No   Requests for exclusive access to clean cache line
PAPI_CA_INV  0x8000000c  Yes   No   Requests for cache line invalidation
PAPI_CA_ITV  0x8000000d  Yes   No   Requests for cache line intervention
```

PAPI: Cache and Memory Hierarchy

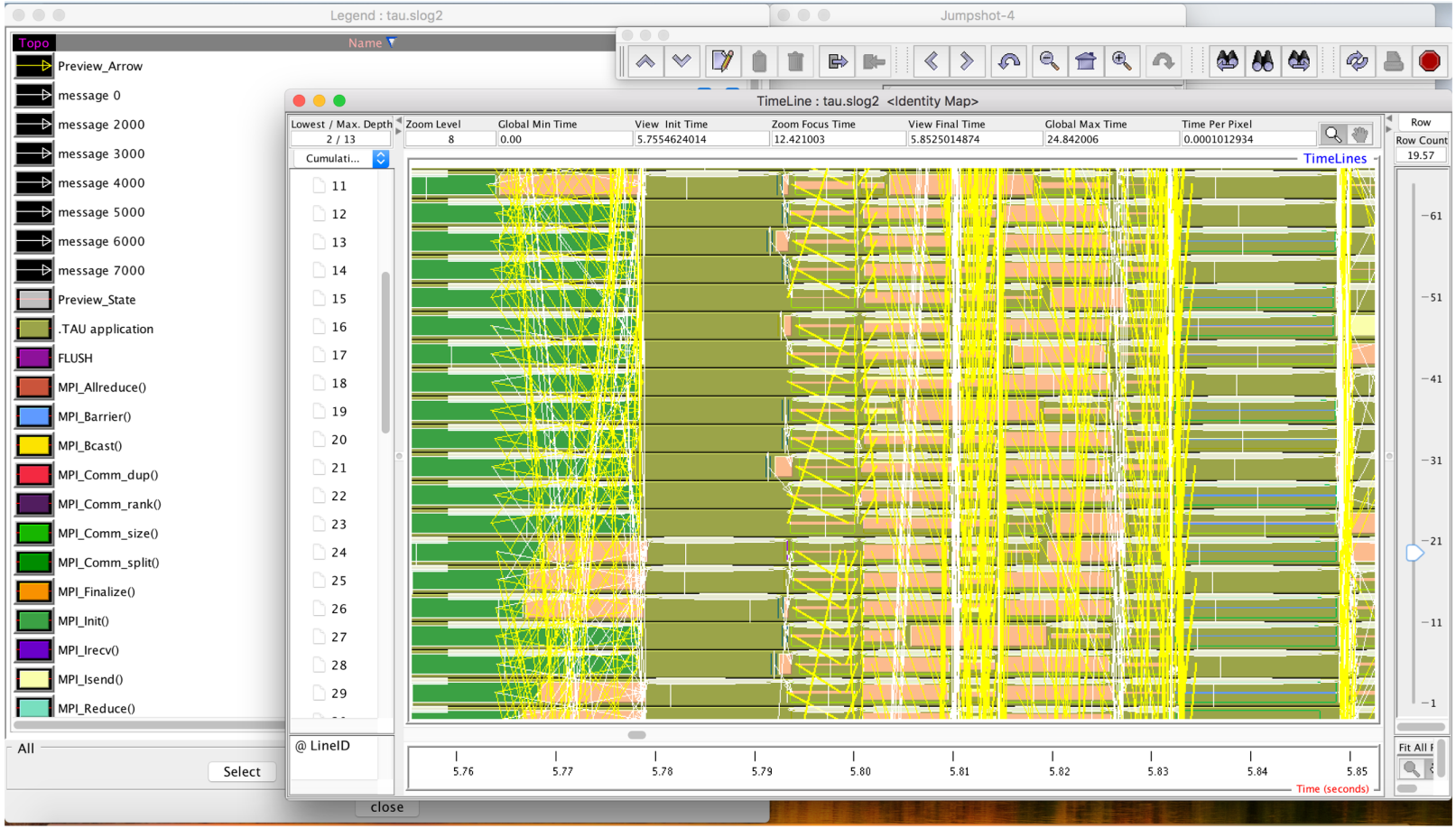
METRIC	Derivation Formula
Graduated loads & stores per cycle	PAPI_LST_INS/PAPI_TOT_CYC
Graduated loads & stores per floating point instruction	PAPI_LST_INS/PAPI_FP_INS
L1 cache line reuse (data)	((PAPI_LST_INS - PAPI_L1_DCM) / PAPI_L1_DCM)
L1 cache data hit rate	1.0 - (PAPI_L1_DCM/PAPI_LST_INS)
L1 data cache read miss ratio	PAPI_L1_DCM/PAPI_L1_DCA
L2 cache line reuse (data)	((PAPI_L1_DCM - PAPI_L2_DCM) / PAPI_L2_DCM)
L2 cache data hit rate	1.0 - (PAPI_L2_DCM/PAPI_L1_DCM)
L2 cache miss ratio	PAPI_L2_TCM/PAPI_L2_TCA
L3 cache line reuse (data)	((PAPI_L2_DCM - PAPI_L3_DCM) / PAPI_L3_DCM)
L3 cache data hit rate	1.0 - (PAPI_L3_DCM/PAPI_L2_DCM)
L3 data cache miss ratio	PAPI_L3_DCM/PAPI_L3_DCA
L3 cache data read ratio	PAPI_L3_DCR/PAPI_L3_DCA
L3 cache instruction miss ratio	PAPI_L3_ICM/PAPI_L3_ICR
Bandwidth used (*Lx* cache)	((PAPI_*Lx*_TCM * *Lx_linesize*) / PAPI_TOT_CYC) * *Clock(MHz)*

TAU - Tracing

```
#!/bin/bash
#SBATCH --job-name="example-trace"
#SBATCH --output="example-trace.%j.%N.out"
#SBATCH --partition=compute
#SBATCH --nodes=3
#SBATCH --ntasks-per-node=24
#SBATCH --export=ALL
#SBATCH -t 00:10:00

module load tau
export TAU_TRACE=1
export TAU_TRACK_MESSAGE=1
export TAU_COMM_MATRIX=1
export SLURM_NODEFILE=`generate_pbs_nodefile`
cd /oasis/scratch/comet/$USER/temp_project
mkdir TMP1
cd TMP1
mpirun_rsh -export-all -hostfile $SLURM_NODEFILE -np 64 tau_exec -T mpi $SLURM_SUBMIT_DIR/sp.C.64
```

- *tau_treemerge.pl*
- *tau2slog2 tau.trc tau.edf -o tau.slog2*
- *jumpshot tau.slog2*

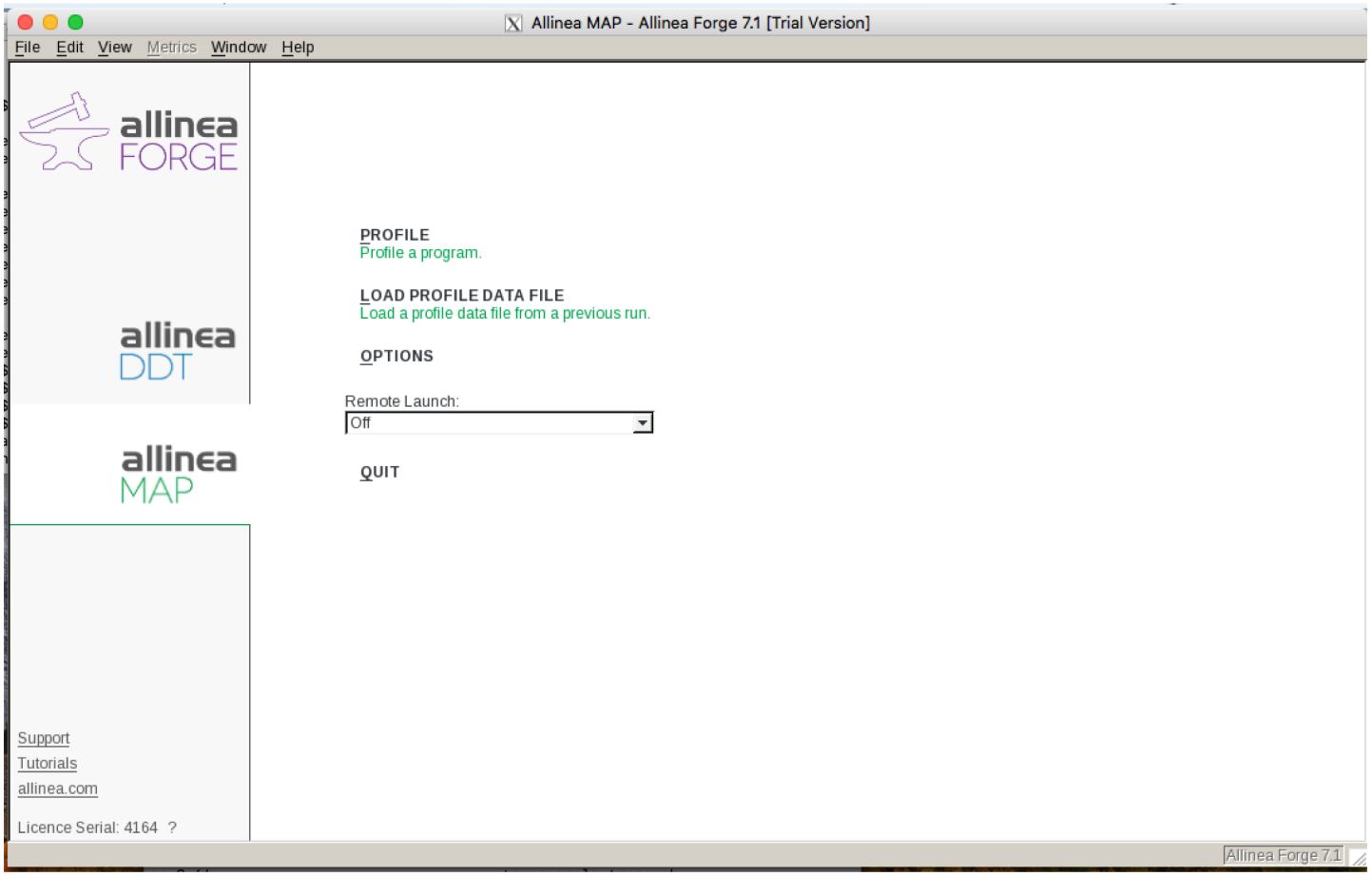
TAU – Jumpshot



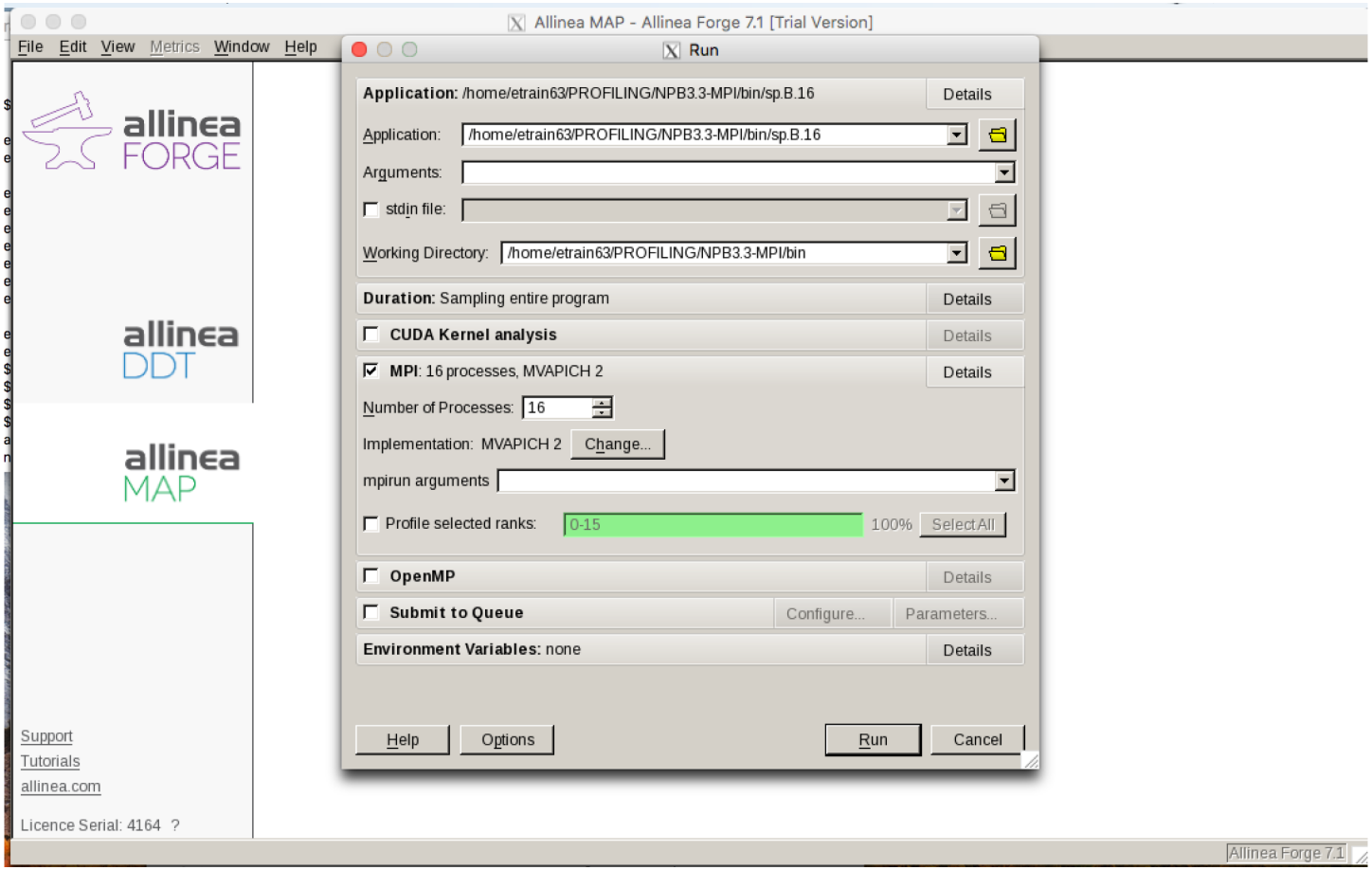
Allinea MAP

- Commercial solution
- Highly scalable, can analyze and visually display the performance
- Processor instruction level – track memory bandwidth bottlenecks, floating-point vectorization
- Memory usage
- I/O
- Communication - such as MPI
- Threads
- MPI, OpenMP, CUDA, Hybrid

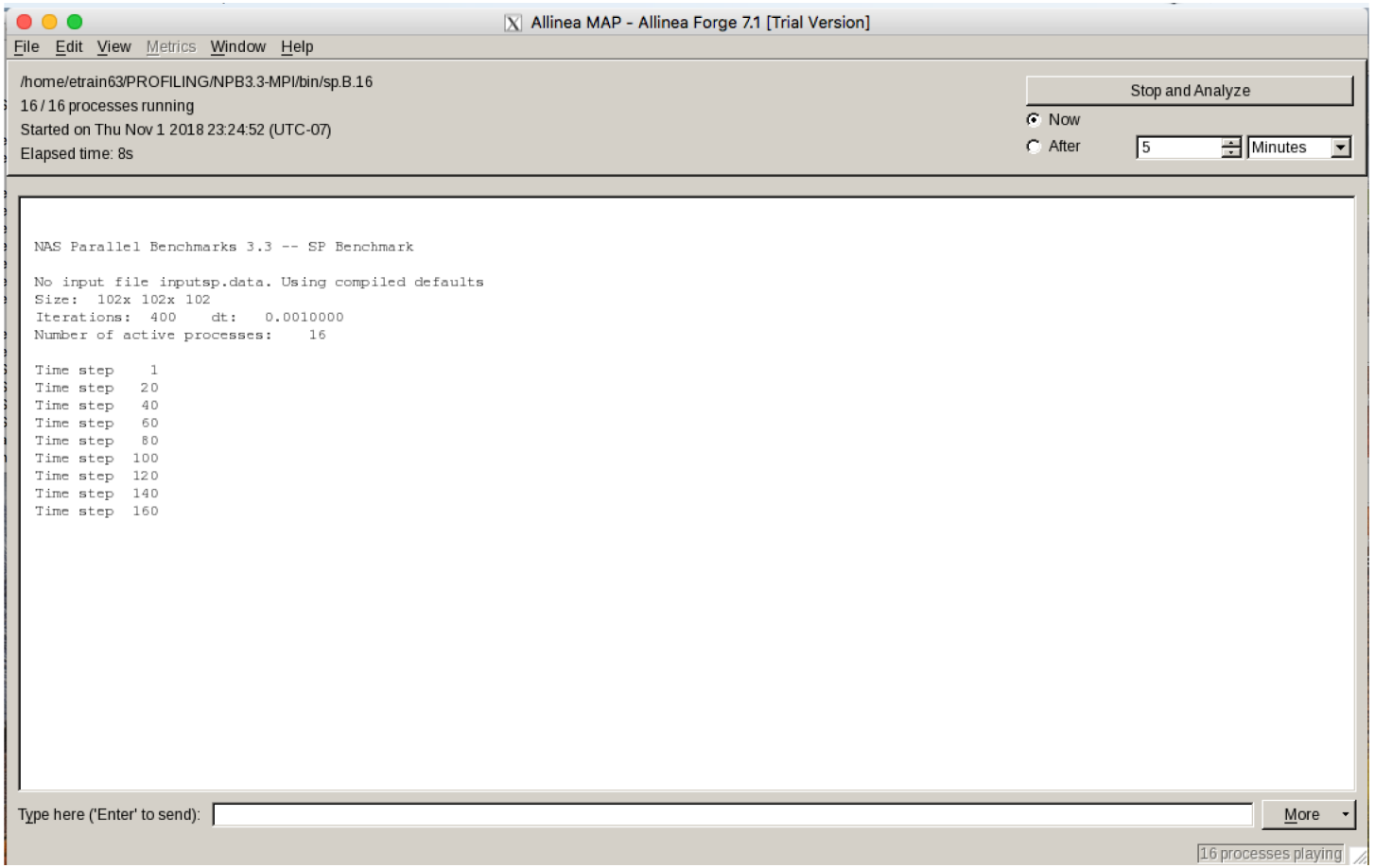
Allinea MAP


Allinea MAP - Allinea Forge 7.1 [Trial Version]
File Edit View Metrics Window Help
allinea FORGE
allinea DDT
allinea MAP
PROFILE
Profile a program.
LOAD PROFILE DATA FILE
Load a profile data file from a previous run.
OPTIONS
Remote Launch:
Off
QUIT
Support
Tutorials
allinea.com
Licence Serial: 4164 ?
Allinea Forge 7.1

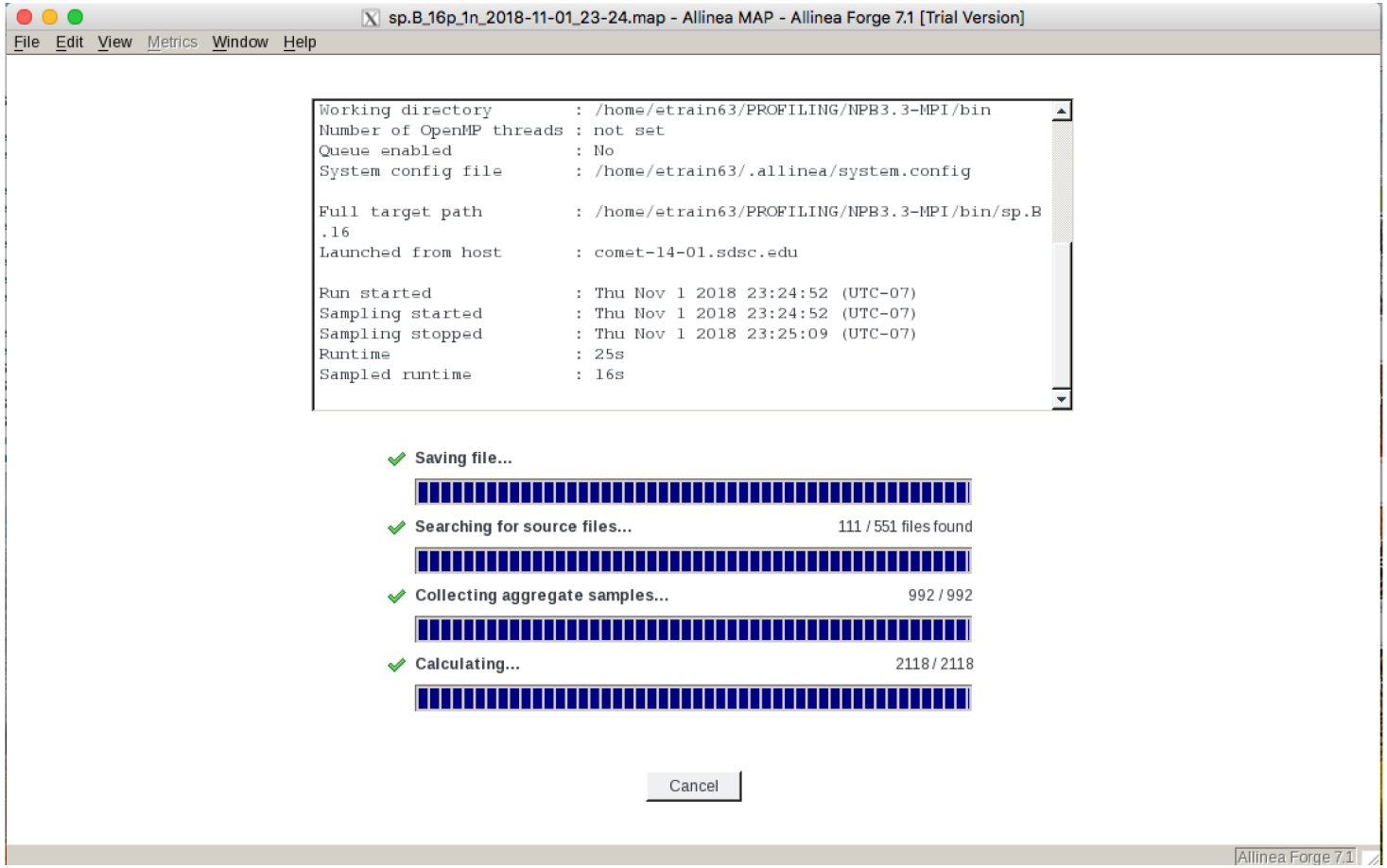
Allinea MAP


Allinea MAP - Allinea Forge 7.1 [Trial Version]
File Edit View Metrics Window Help
allinea FORGE
allinea DDT
allinea MAP
Support
Tutorials
allinea.com
Licence Serial: 4164 ?
Run
Application: /home/etrain63/PROFILING/NPB3.3-MPI/bin/sp.B.16
Details
Application: /home/etrain63/PROFILING/NPB3.3-MPI/bin/sp.B.16
Arguments:
stdin file:
Working Directory: /home/etrain63/PROFILING/NPB3.3-MPI/bin
Duration: Sampling entire program
Details
CUDA Kernel analysis
Details
MPI: 16 processes, MVAPICH 2
Details
Number of Processes: 16
Implementation: MVAPICH 2
Change...
mpirun arguments
Profile selected ranks:
0-15
100%
SelectAll
OpenMP
Details
Submit to Queue
Configure...
Parameters...
Environment Variables: none
Details
Help
Options
Run
Cancel
Allinea Forge 7.1

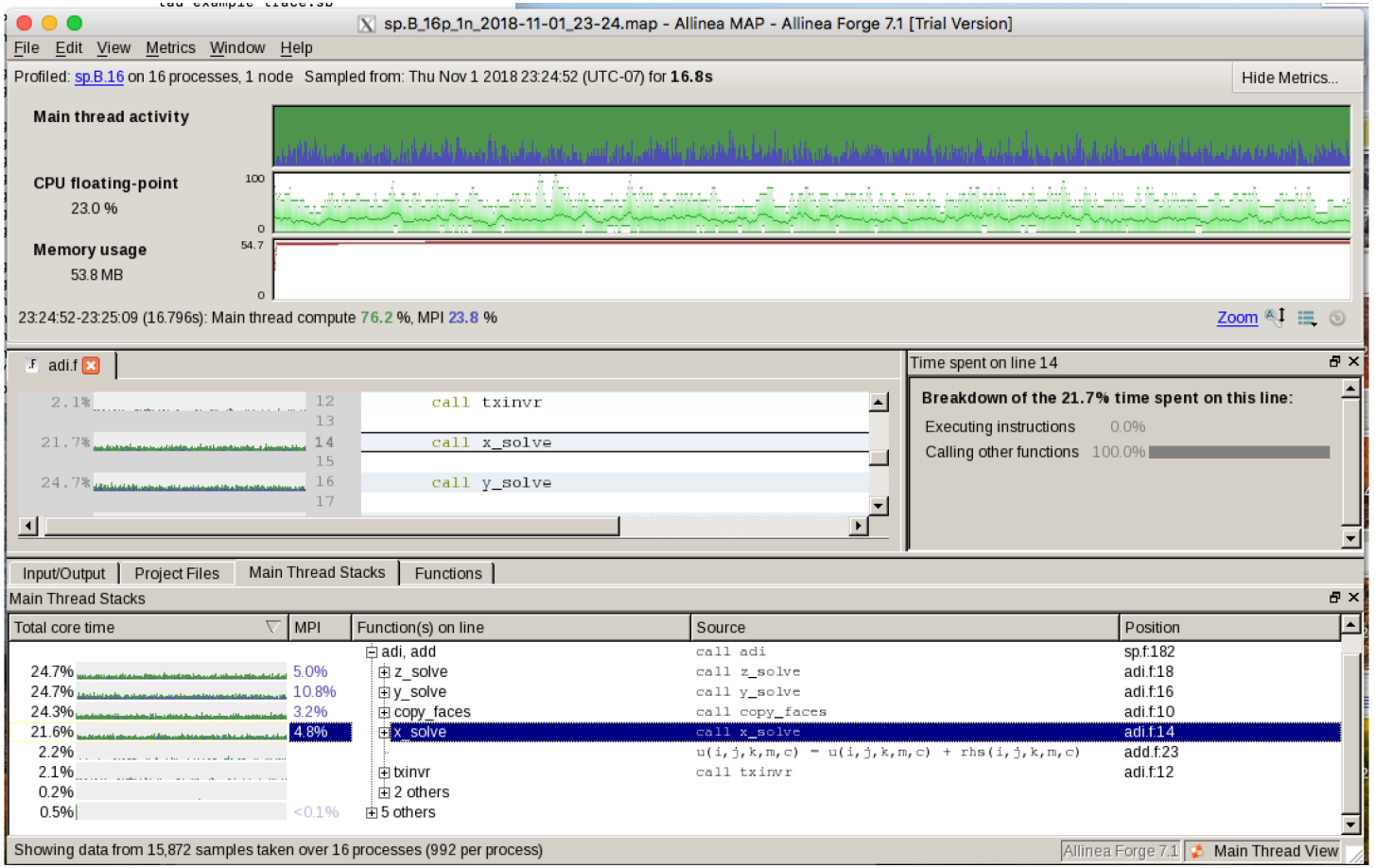
Allinea MAP


Allinea MAP - Allinea Forge 7.1 [Trial Version]
File Edit View Metrics Window Help
/home/etrain63/PROFILING/NPB3.3-MPI/bin/sp.B.16
16 / 16 processes running
Started on Thu Nov 1 2018 23:24:52 (UTC-07)
Elapsed time: 8s
Stop and Analyze
Now
After
5
Minutes
NAS Parallel Benchmarks 3.3 -- SP Benchmark
No input file inputsp.data. Using compiled defaults
Size: 102x 102x 102
Iterations: 400 dt: 0.0010000
Number of active processes: 16
Time step 1
Time step 20
Time step 40
Time step 60
Time step 80
Time step 100
Time step 120
Time step 140
Time step 160
Type here ('Enter' to send):
More
16 processes playing

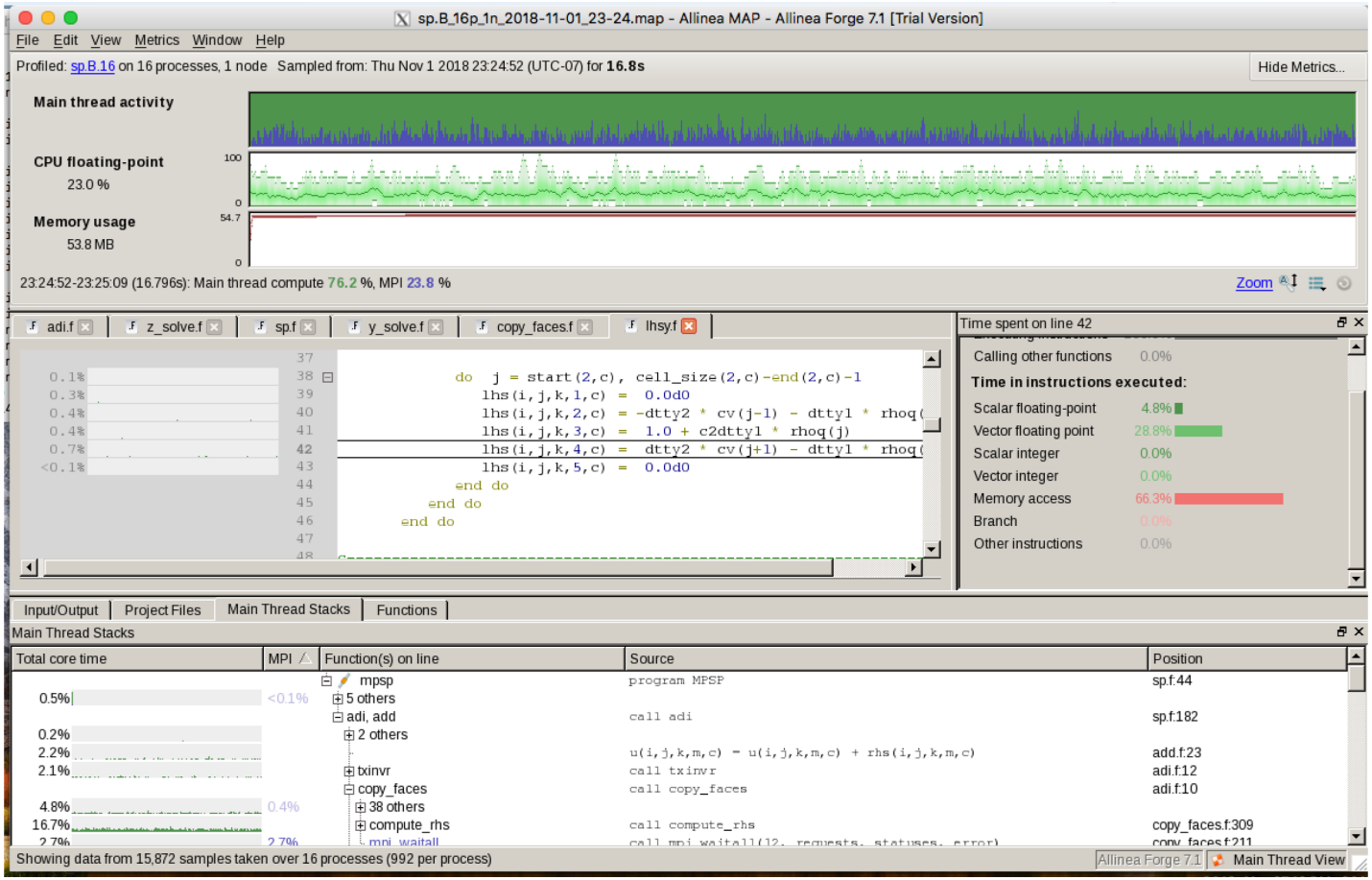
Allinea MAP


sp.B_16p_1n_2018-11-01_23-24.map - Allinea MAP - Allinea Forge 7.1 [Trial Version]
File Edit View Metrics Window Help
Working directory : /home/etrain63/PROFILING/NPB3.3-MPI/bin
Number of OpenMP threads : not set
Queue enabled : No
System config file : /home/etrain63/.allinea/system.config
Full target path : /home/etrain63/PROFILING/NPB3.3-MPI/bin/sp.B.16
Launched from host : comet-14-01.sdsc.edu
Run started : Thu Nov 1 2018 23:24:52 (UTC-07)
Sampling started : Thu Nov 1 2018 23:24:52 (UTC-07)
Sampling stopped : Thu Nov 1 2018 23:25:09 (UTC-07)
Runtime : 25s
Sampled runtime : 16s
Saving file...
Searching for source files... 111 / 551 files found
Collecting aggregate samples... 992 / 992
Calculating... 2118 / 2118
Cancel
Allinea Forge 7.1

Allinea MAP


sp.B_16p_1n_2018-11-01_23-24.map - Allinea MAP - Allinea Forge 7.1 [Trial Version]
File Edit View Metrics Window Help
Profiled: sp.B.16 on 16 processes, 1 node Sampled from: Thu Nov 1 2018 23:24:52 (UTC-07) for 16.8s
Hide Metrics...
Main thread activity
CPU floating-point
23.0 %
100
0
Memory usage
53.8 MB
54.7
0
23:24:52-23:25:09 (16.796s): Main thread compute 76.2 %, MPI 23.8 %
Zoom
adi.f
2.1%
21.7%
24.7%
12 call txinvr
13
14 call x_solve
15
16 call y_solve
17
Time spent on line 14
Breakdown of the 21.7% time spent on this line:
Executing instructions 0.0%
Calling other functions 100.0%
Input/Output
Project Files
Main Thread Stacks
Functions
Main Thread Stacks
Total core time
MPI
Function(s) on line
Source
Position
adi, add call adi sp.f:182
24.7% 5.0% z_solve call z_solve adi.f:18
24.7% 10.8% y_solve call y_solve adi.f:16
24.3% 3.2% copy_faces call copy_faces adi.f:10
21.6% 4.8% x_solve call x_solve adi.f:14
2.2% u(i,j,k,m,c) = u(i,j,k,m,c) + rhs(i,j,k,m,c) add.f:23
2.1% txinvr call txinvr adi.f:12
0.2% 2 others
0.5% <0.1% 5 others
Showing data from 15,872 samples taken over 16 processes (992 per process)
Allinea Forge 7.1
Main Thread View

Allinea MAP


sp.B_16p_1n_2018-11-01_23-24.map - Allinea MAP - Allinea Forge 7.1 [Trial Version]
File Edit View Metrics Window Help
Profiled: sp.B.16 on 16 processes, 1 node Sampled from: Thu Nov 1 2018 23:24:52 (UTC-07) for 16.8s
Hide Metrics...
Main thread activity
CPU floating-point
23.0 %
Memory usage
53.8 MB
23:24:52-23:25:09 (16.796s): Main thread compute 76.2 %, MPI 23.8 %
Zoom
adi.f z_solve.f sp.f y_solve.f copy_faces.f lhsy.f
do j = start(2,c), cell_size(2,c)-end(2,c)-1
lhs(i,j,k,1,c) = 0.0d0
lhs(i,j,k,2,c) = -dtty2 * cv(j-1) - dtty1 * rhoq(
lhs(i,j,k,3,c) = 1.0 + c2dtty1 * rhoq(j)
lhs(i,j,k,4,c) = dtty2 * cv(j+1) - dtty1 * rhoq(
lhs(i,j,k,5,c) = 0.0d0
end do
end do
end do
Time spent on line 42
Calling other functions 0.0%
Time in instructions executed:
Scalar floating-point 4.8%
Vector floating point 28.8%
Scalar integer 0.0%
Vector integer 0.0%
Memory access 66.3%
Branch 0.0%
Other instructions 0.0%
Input/Output Project Files Main Thread Stacks Functions
Main Thread Stacks
Total core time MPI Function(s) on line Source Position
mpsp program MPSP sp.f:44
5 others
adi, add call adi sp.f:182
2 others
u(i,j,k,m,c) = u(i,j,k,m,c) + rhs(i,j,k,m,c) add.f:23
txinvr call txinvr adi.f:12
copy_faces call copy_faces adi.f:10
38 others
compute_rhs call compute_rhs copy_faces.f:309
Showing data from 15,872 samples taken over 16 processes (992 per process)
Allinea Forge 7.1
Main Thread View

Summary

- Variety of profiling/tracing tools – mpiP, TAU, Allinea MAP (and others)
- For MPI jobs the simplest lightweight option is mpiP.
- TAU has both profiling and tracing options, can get data from hardware counters (uses PAPI).
- Allinea MAP takes a sampling based approach to minimize overhead. GUI based, easy to drill into details