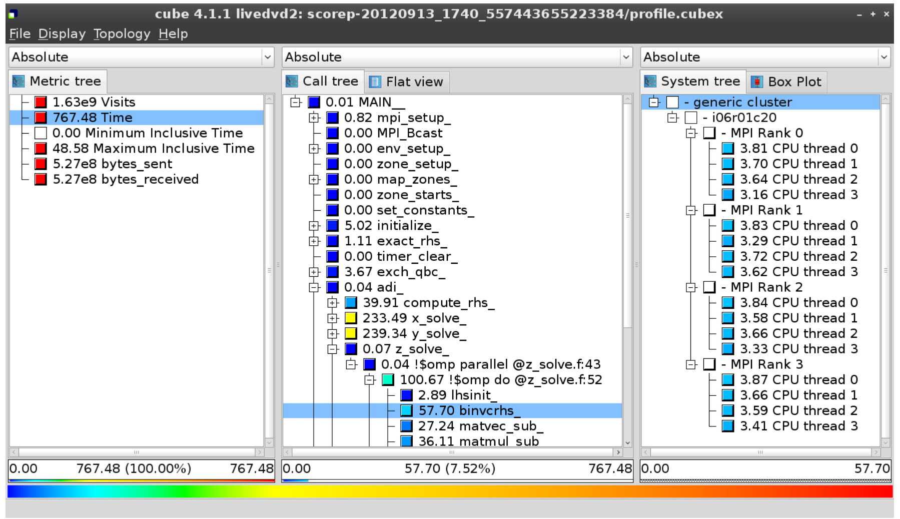
Analysis report examination with CUBE

The Scalasca Team
Jülich Supercomputing Centre

CUBE

CubeLib DOI 10.5281/zenodo.7737408
CubeGUI DOI 10.5281/zenodo.7737411

- Parallel program analysis report exploration tools
  - Libraries for XML+binary report reading & writing
  - Algebra utilities for report processing
  - GUI for interactive analysis exploration
    - Requires Qt ≥ 5



```
cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex
File  Display  Topology  Help

Absolute                          Absolute                               Absolute
Metric tree                       Call tree | Flat view                  System tree | Box Plot
1.63e9 Visits                     0.01 MAIN__                            - generic cluster
767.48 Time                         0.82 mpi_setup_                        - i06r01c20
0.00 Minimum Inclusive Time         0.00 MPI_Bcast                           - MPI Rank 0
48.58 Maximum Inclusive Time        0.00 env_setup_                            3.81 CPU thread 0
5.27e8 bytes_sent                   0.00 zone_setup_                           3.70 CPU thread 1
5.27e8 bytes_received               0.00 map_zones_                            3.64 CPU thread 2
                                    0.00 zone_starts_                          3.16 CPU thread 3
                                    0.00 set_constants_                      - MPI Rank 1
                                    5.02 initialize_                           3.83 CPU thread 0
                                    1.11 exact_rhs_                            3.29 CPU thread 1
                                    0.00 timer_clear_                          3.72 CPU thread 2
                                    3.67 exch_qbc_                             3.62 CPU thread 3
                                    0.04 adi_                                - MPI Rank 2
                                      39.91 compute_rhs_                       3.84 CPU thread 0
                                      233.49 x_solve_                          3.58 CPU thread 1
                                      239.34 y_solve_                          3.66 CPU thread 2
                                      0.07 z_solve_                            3.33 CPU thread 3
                                        0.04 !$omp parallel @z_solve.f:43    - MPI Rank 3
                                          100.67 !$omp do @z_solve.f:52        3.87 CPU thread 0
                                            2.89 lhsinit_                      3.66 CPU thread 1
                                            57.70 binvcrhs_                    3.59 CPU thread 2
                                            27.24 matvec_sub_                  3.41 CPU thread 3
                                            36.11 matmul_sub_
0.00   767.48 (100.00%)   767.48  0.00   57.70 (7.52%)   767.48      0.00                57.70
```

- Originally developed as part of the Scalasca toolset

- Now available as a separate component
  - Can be installed independently of Score-P, e.g., on laptop or desktop
  - Latest release: Cube v4.8.2 (Oct 2023)

Note: source distribution tarballs for Linux, as well as binary packages provided for Windows & MacOS, from **www.scalasca.org** website in software/Cube-4x

Cube GUI (karolina)

mailto: **scalasca@fz-juelich.de**

cube scalasca

- Run ***remote*** *(often convenient)*
  - start X server (e.g., Xming) locally, or use alternative such as mobaXterm
  - connect to Karolina with X forwarding enabled
    - **-Y** may be faster but is insecure!
  - load module and start cube remotely

```
desk$ ssh -X login.karolina.it4i.cz
Welcome to Karolina…
karolina$ module load CubeGUI
karolina$ cube ./scorep_sum/profile.cubex
```

Sample measurements (CUBE files) on Karolina:
/mnt/proj2/dd-24-88/jsc/samples

- Install & run ***local*** *(recommended)*
  - install Cube GUI locally on desktop
    - binary packages available for MacOS & Windows and externally provided by OpenHPC and various Linux distributions
    - source package available for Linux, requires Qt
      - configure/build/install manually or use your favourite framework (e.g. Spack or EasyBuild)
  - copy .cubex file (or entire scorep directory) to desktop from remote system
    OR locally mount remote filesystem
  - start cube locally

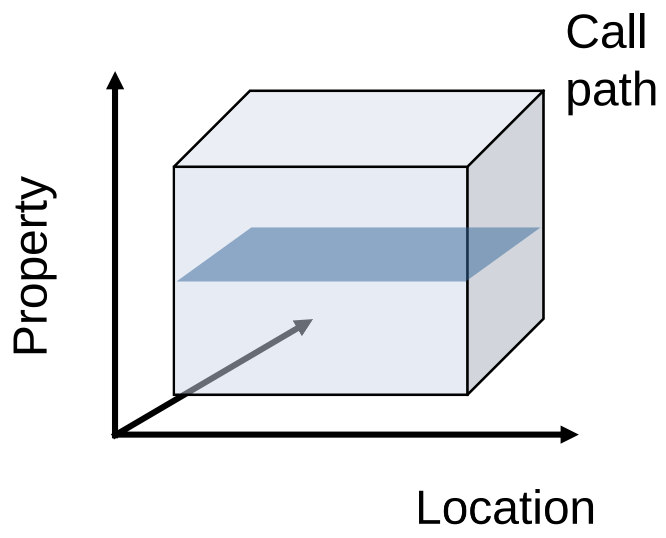
```
desk$ mkdir $HOME/mnt
desk$ sshfs [user@]remote.sys:[dir] $HOME/mnt
desk$ cd $HOME/mnt
desk$ cube ./scorep_sum/profile.cubex
```

Analysis presentation and exploration

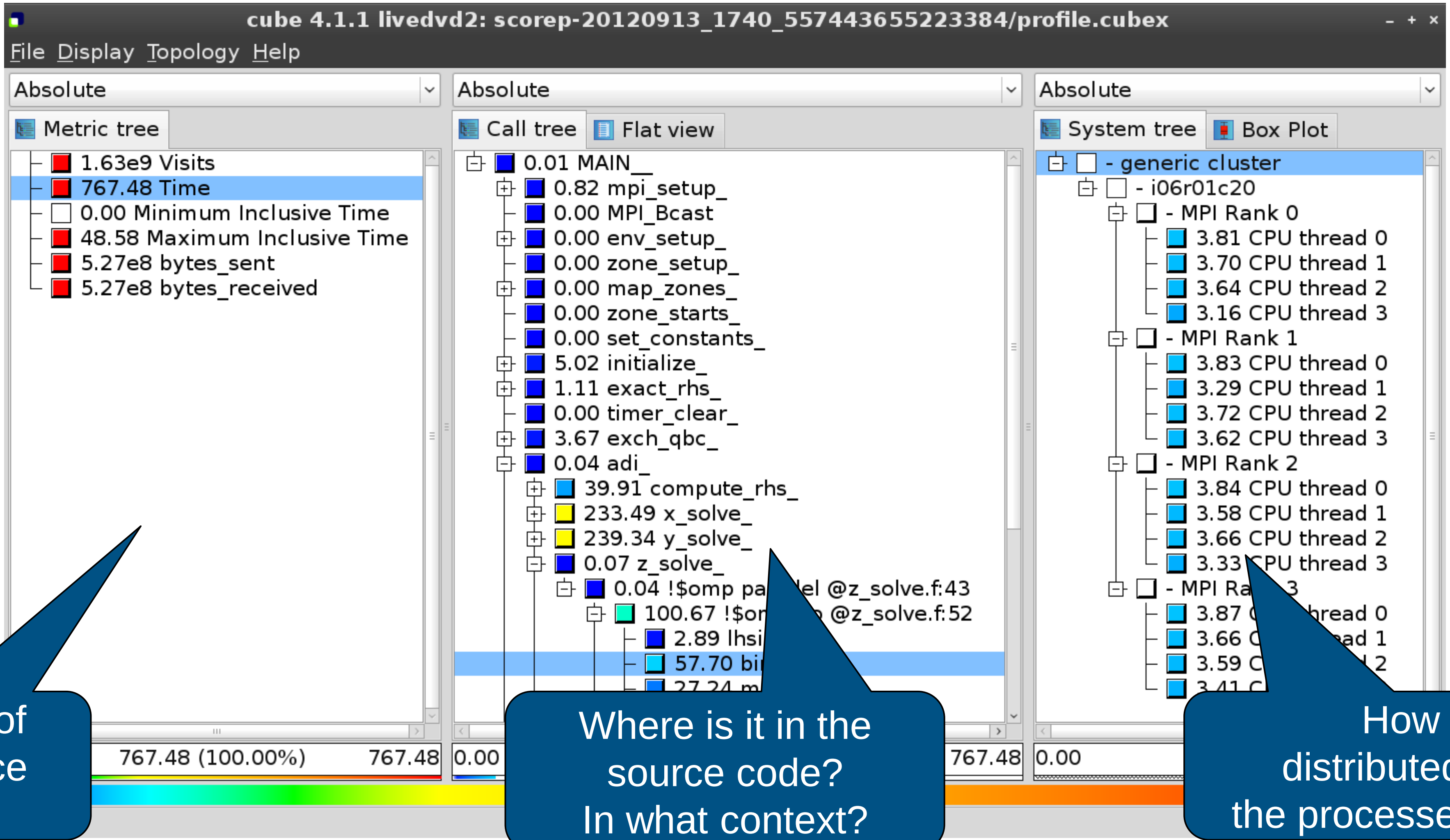
- Representation of values (severity matrix) on three hierarchical axes
  - Performance property (metric)
  - Call path (program location)
  - System location (process/thread/stream)

- Three coupled tree browsers

- Cube displays severities
  - As *value*: for precise comparison
  - As *colour*: for easy identification of hotspots
  - *Inclusive* value when closed & *exclusive* value when expanded
  - Customizable via display modes



Analysis presentation


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex
File Display Topology Help
Absolute
Metric tree
Call tree
Flat view
System tree
Box Plot

```
1.63e9 Visits
767.48 Time
0.00 Minimum Inclusive Time
48.58 Maximum Inclusive Time
5.27e8 bytes_sent
5.27e8 bytes_received
```

```
0.01 MAIN__
  0.82 mpi_setup_
  0.00 MPI_Bcast
  0.00 env_setup_
  0.00 zone_setup_
  0.00 map_zones_
  0.00 zone_starts_
  0.00 set_constants_
  5.02 initialize_
  1.11 exact_rhs_
  0.00 timer_clear_
  3.67 exch_qbc_
  0.04 adi_
    39.91 compute_rhs_
    233.49 x_solve_
    239.34 y_solve_
    0.07 z_solve_
      0.04 !$omp pa    lel @z_solve.f:43
        100.67 !$om      @z_solve.f:52
          2.89 lhsi
          57.70 bi
          27.24 m
```

```
- generic cluster
  - i06r01c20
    - MPI Rank 0
      3.81 CPU thread 0
      3.70 CPU thread 1
      3.64 CPU thread 2
      3.16 CPU thread 3
    - MPI Rank 1
      3.83 CPU thread 0
      3.29 CPU thread 1
      3.72 CPU thread 2
      3.62 CPU thread 3
    - MPI Rank 2
      3.84 CPU thread 0
      3.58 CPU thread 1
      3.66 CPU thread 2
      3.33 CPU thread 3
    - MPI Ra    3
      3.87 C    hread 0
      3.66 C    ead 1
      3.59 C      2
      3.41 C
```

767.48 (100.00%) 767.48 0.00 767.48 0.00

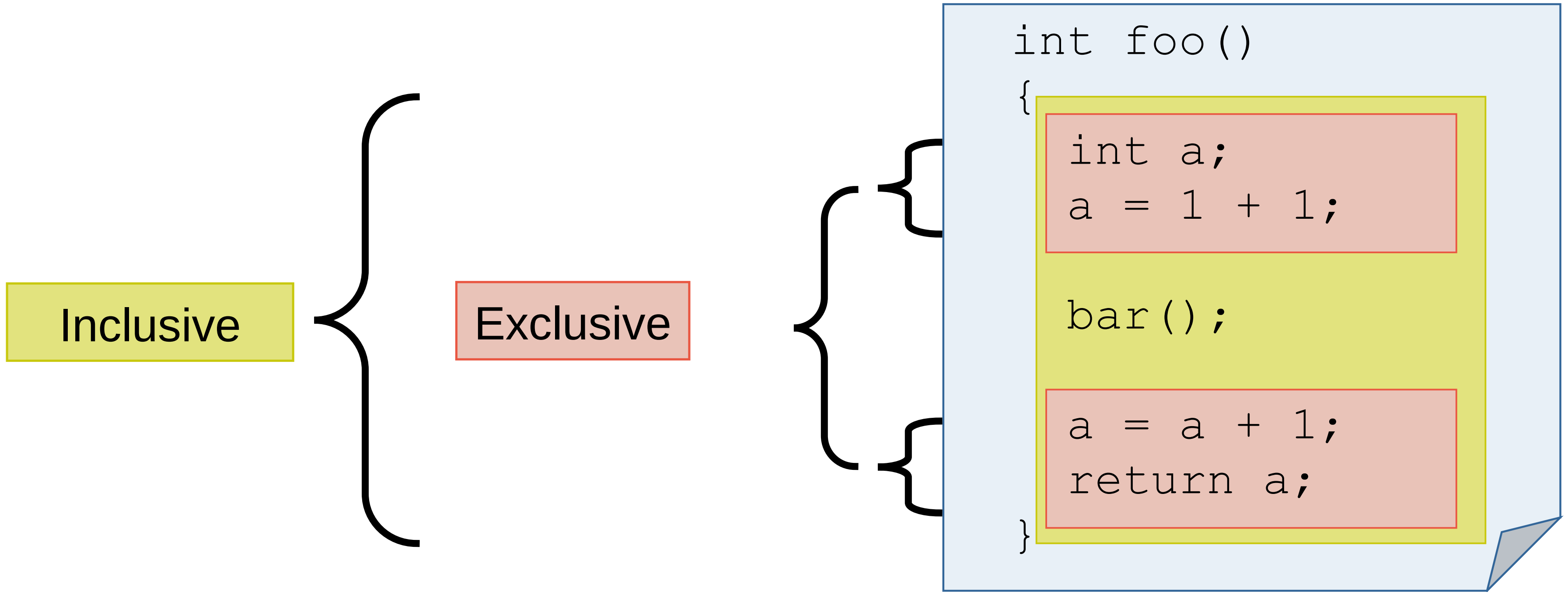
What kind of performance metric?

Where is it in the source code? In what context?

How is it distributed across the processes/threads?

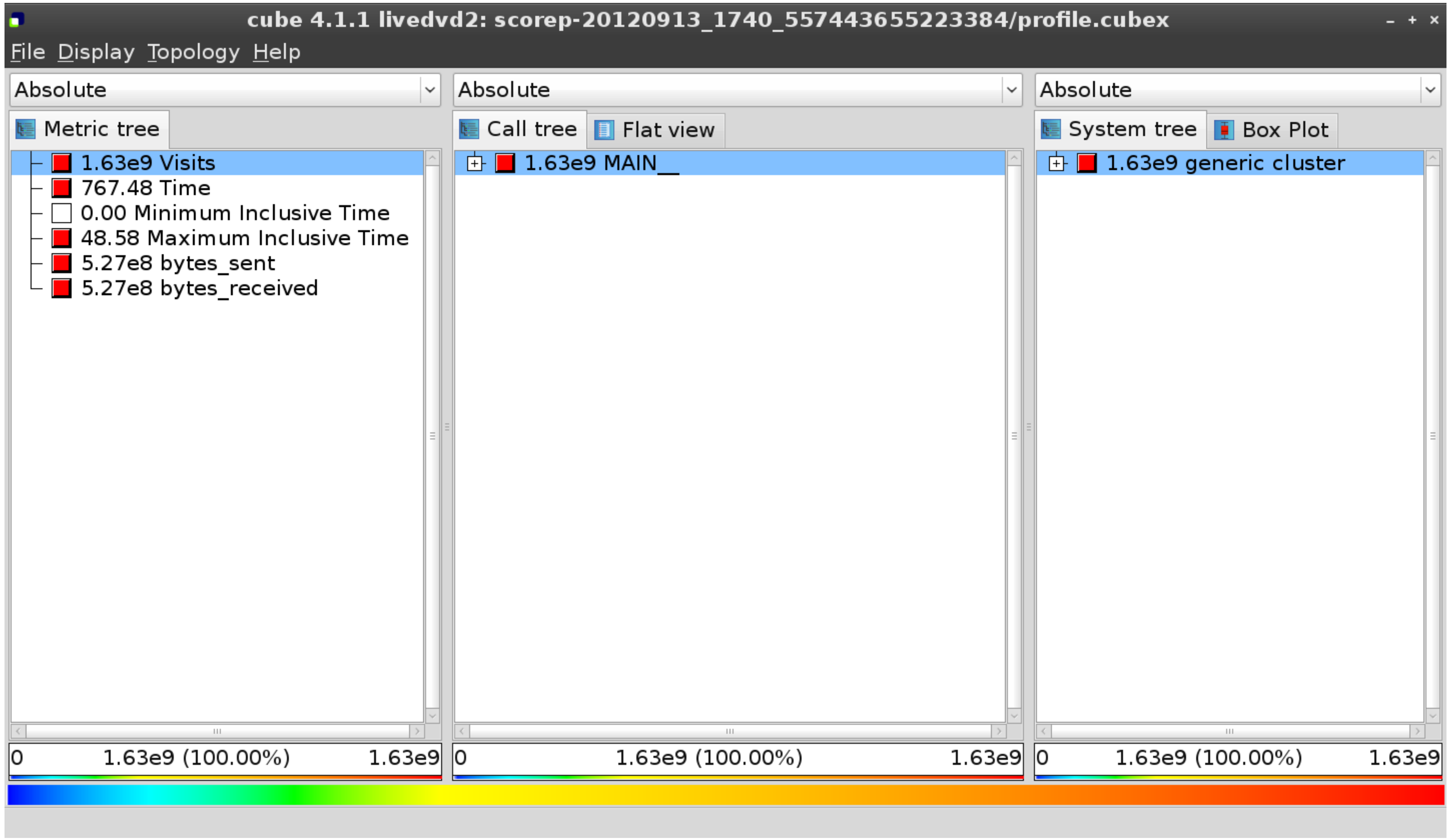
Inclusive vs. exclusive values

- Inclusive
  - Information of all sub-elements aggregated into single value
- Exclusive
  - Information cannot be subdivided further



```
int foo()
{
  int a;
  a = 1 + 1;

  bar();

  a = a + 1;
  return a;
}
```

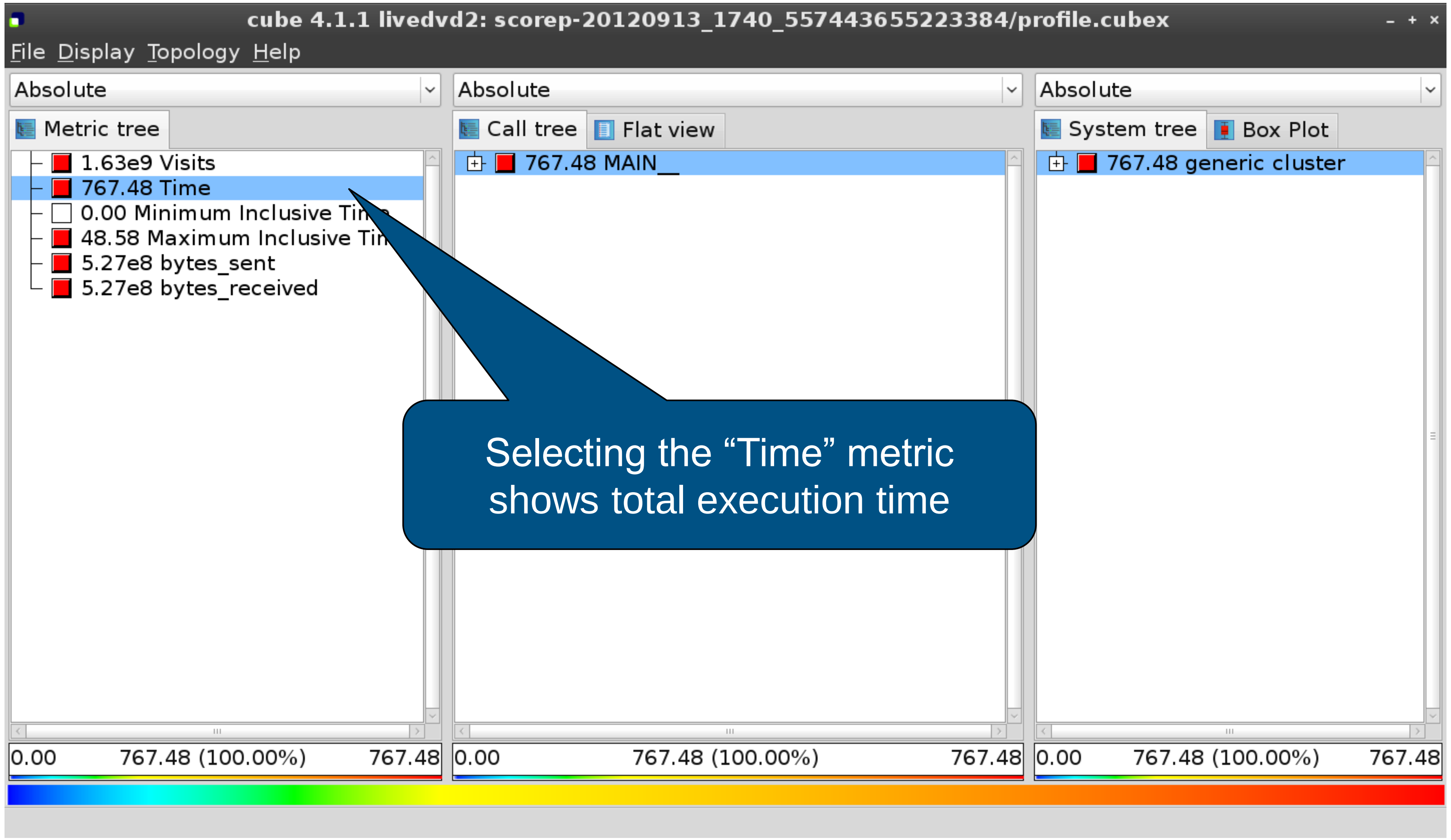
Score-P analysis report exploration (opening view)



cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

```
Metric tree:
├ 1.63e9 Visits
├ 767.48 Time
├ 0.00 Minimum Inclusive Time
├ 48.58 Maximum Inclusive Time
├ 5.27e8 bytes_sent
└ 5.27e8 bytes_received

Call tree:
⊞ 1.63e9 MAIN__

System tree:
⊞ 1.63e9 generic cluster

0    1.63e9 (100.00%)    1.63e9
0    1.63e9 (100.00%)    1.63e9
0    1.63e9 (100.00%)    1.63e9
```

Metric selection


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex
File Display Topology Help
Absolute
Absolute
Absolute
Metric tree
Call tree
Flat view
System tree
Box Plot

```
1.63e9 Visits
767.48 Time
0.00 Minimum Inclusive Time
48.58 Maximum Inclusive Time
5.27e8 bytes_sent
5.27e8 bytes_received
```

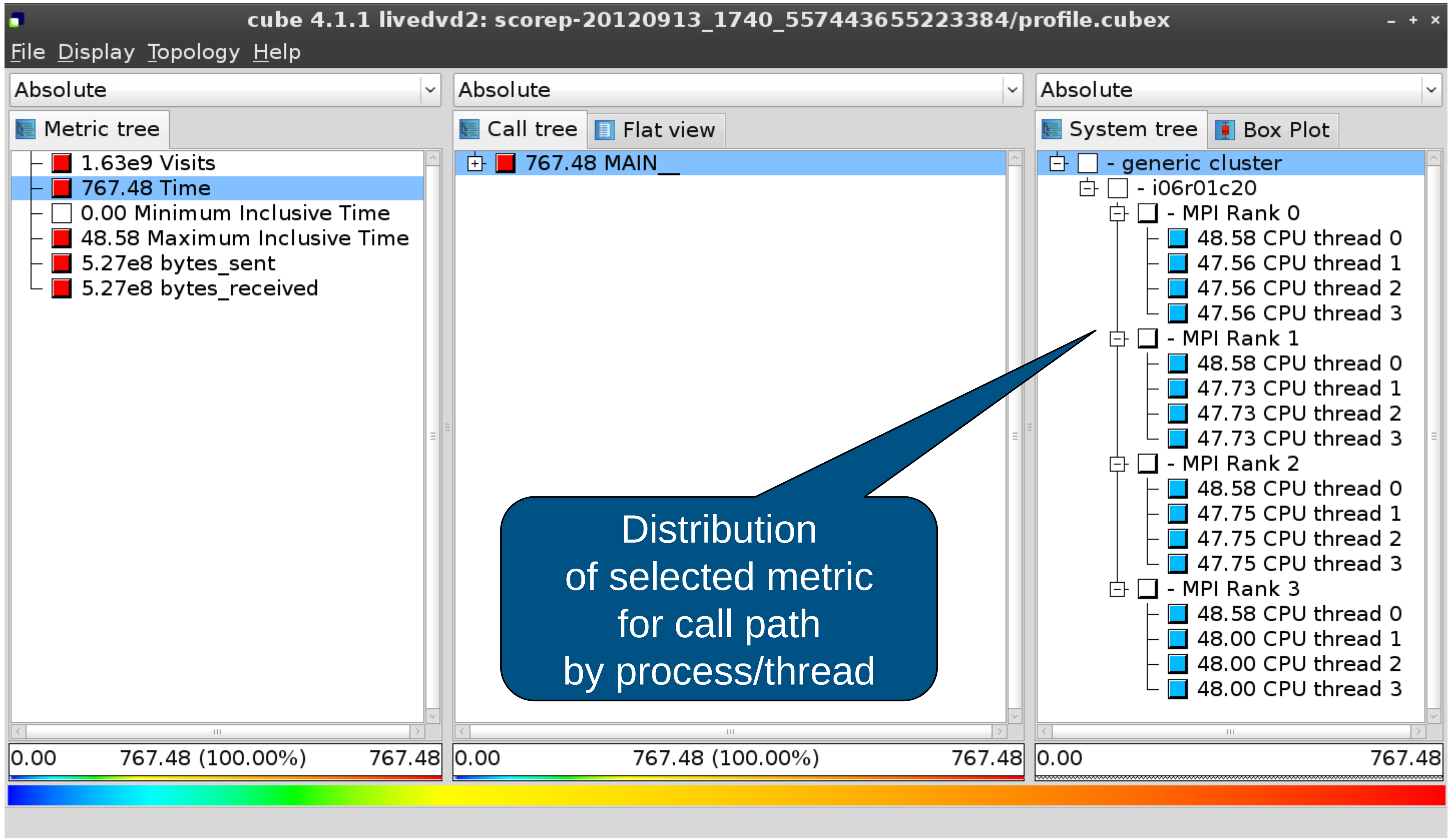
```
767.48 MAIN__
```

```
767.48 generic cluster
```

```
0.00    767.48 (100.00%)    767.48
```

Selecting the "Time" metric shows total execution time

Expanding the system tree



cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

File Display Topology Help

Absolute | Absolute | Absolute

Metric tree
- 1.63e9 Visits
- 767.48 Time
- 0.00 Minimum Inclusive Time
- 48.58 Maximum Inclusive Time
- 5.27e8 bytes_sent
- 5.27e8 bytes_received

Call tree | Flat view
- 767.48 MAIN__

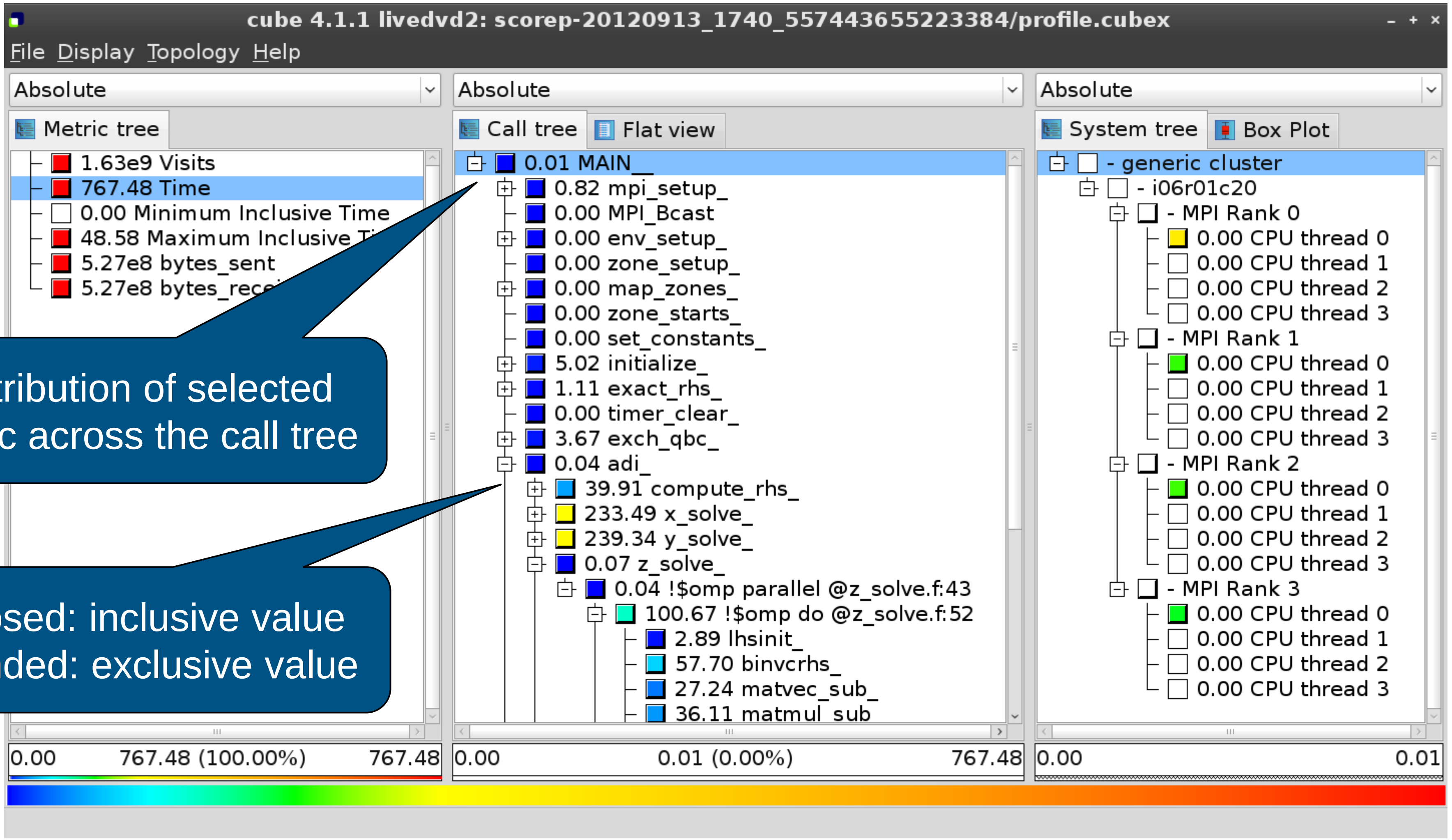
System tree | Box Plot

```
- generic cluster
  - i06r01c20
    - MPI Rank 0
      48.58 CPU thread 0
      47.56 CPU thread 1
      47.56 CPU thread 2
      47.56 CPU thread 3
    - MPI Rank 1
      48.58 CPU thread 0
      47.73 CPU thread 1
      47.73 CPU thread 2
      47.73 CPU thread 3
    - MPI Rank 2
      48.58 CPU thread 0
      47.75 CPU thread 1
      47.75 CPU thread 2
      47.75 CPU thread 3
    - MPI Rank 3
      48.58 CPU thread 0
      48.00 CPU thread 1
      48.00 CPU thread 2
      48.00 CPU thread 3
```

0.00 767.48 (100.00%) 767.48 | 0.00 767.48 (100.00%) 767.48 | 0.00 767.48

Distribution of selected metric for call path by process/thread

Expanding the call tree



cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

File Display Topology Help

Absolute | Absolute | Absolute

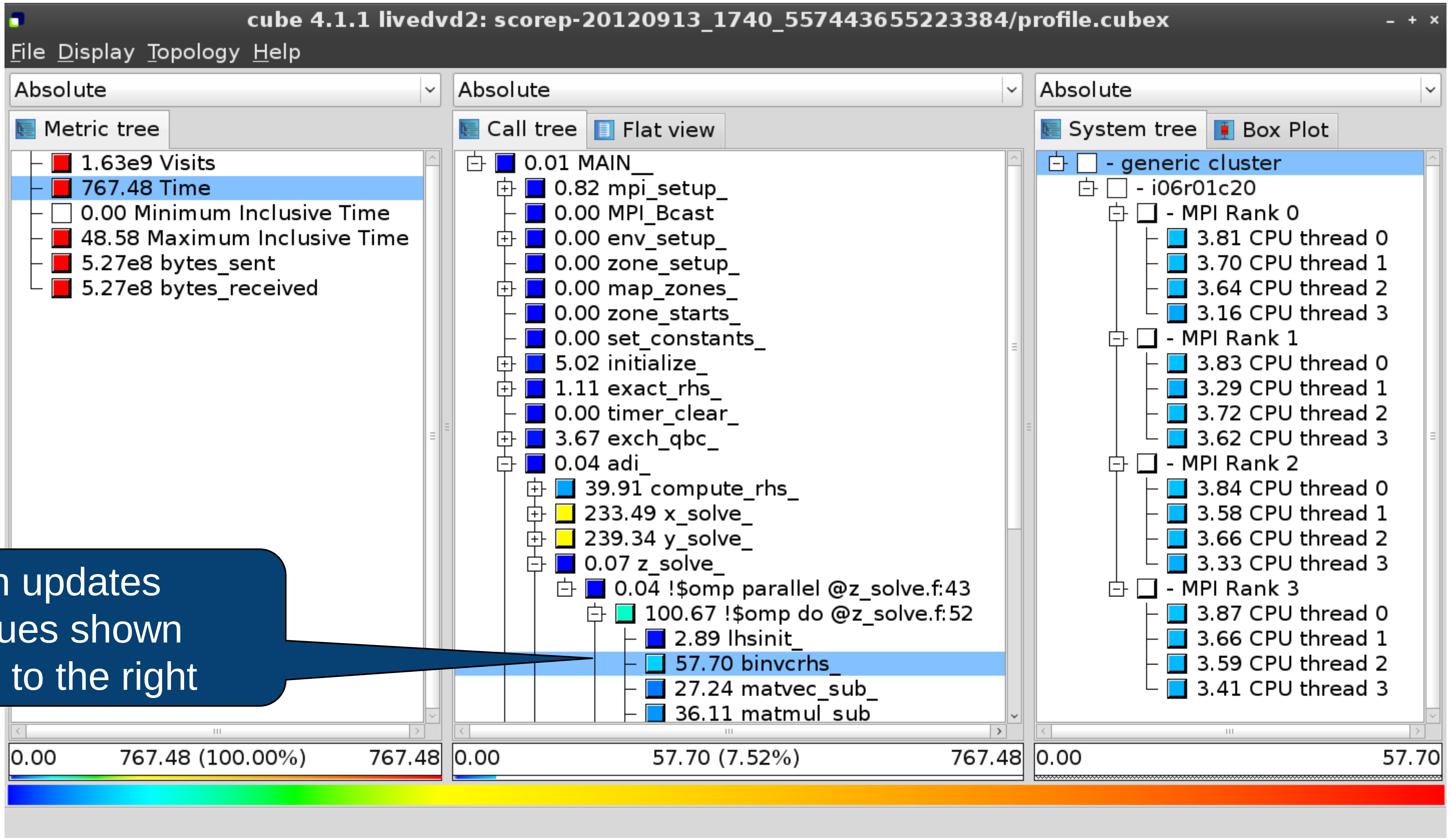
Metric tree | Call tree | Flat view | System tree | Box Plot

```
Metric tree:
├ 1.63e9 Visits
├ 767.48 Time
├ 0.00 Minimum Inclusive Time
├ 48.58 Maximum Inclusive Ti
├ 5.27e8 bytes_sent
└ 5.27e8 bytes_rece

Call tree:
⊟ 0.01 MAIN__
  ⊞ 0.82 mpi_setup_
    0.00 MPI_Bcast
  ⊞ 0.00 env_setup_
    0.00 zone_setup_
  ⊞ 0.00 map_zones_
    0.00 zone_starts_
    0.00 set_constants_
  ⊞ 5.02 initialize_
  ⊞ 1.11 exact_rhs_
    0.00 timer_clear_
  ⊞ 3.67 exch_qbc_
  ⊟ 0.04 adi_
    ⊞ 39.91 compute_rhs_
    ⊞ 233.49 x_solve_
    ⊞ 239.34 y_solve_
    ⊟ 0.07 z_solve_
      ⊟ 0.04 !$omp parallel @z_solve.f:43
        ⊟ 100.67 !$omp do @z_solve.f:52
            2.89 lhsinit_
            57.70 binvcrhs_
            27.24 matvec_sub_
            36.11 matmul_sub

System tree:
⊟ - generic cluster
  ⊟ - i06r01c20
    ⊟ - MPI Rank 0
        0.00 CPU thread 0
        0.00 CPU thread 1
        0.00 CPU thread 2
        0.00 CPU thread 3
    ⊟ - MPI Rank 1
        0.00 CPU thread 0
        0.00 CPU thread 1
        0.00 CPU thread 2
        0.00 CPU thread 3
    ⊟ - MPI Rank 2
        0.00 CPU thread 0
        0.00 CPU thread 1
        0.00 CPU thread 2
        0.00 CPU thread 3
    ⊟ - MPI Rank 3
        0.00 CPU thread 0
        0.00 CPU thread 1
        0.00 CPU thread 2
        0.00 CPU thread 3

0.00   767.48 (100.00%)   767.48 | 0.00   0.01 (0.00%)   767.48 | 0.00   0.01
```

Distribution of selected metric across the call tree

Collapsed: inclusive value
Expanded: exclusive value

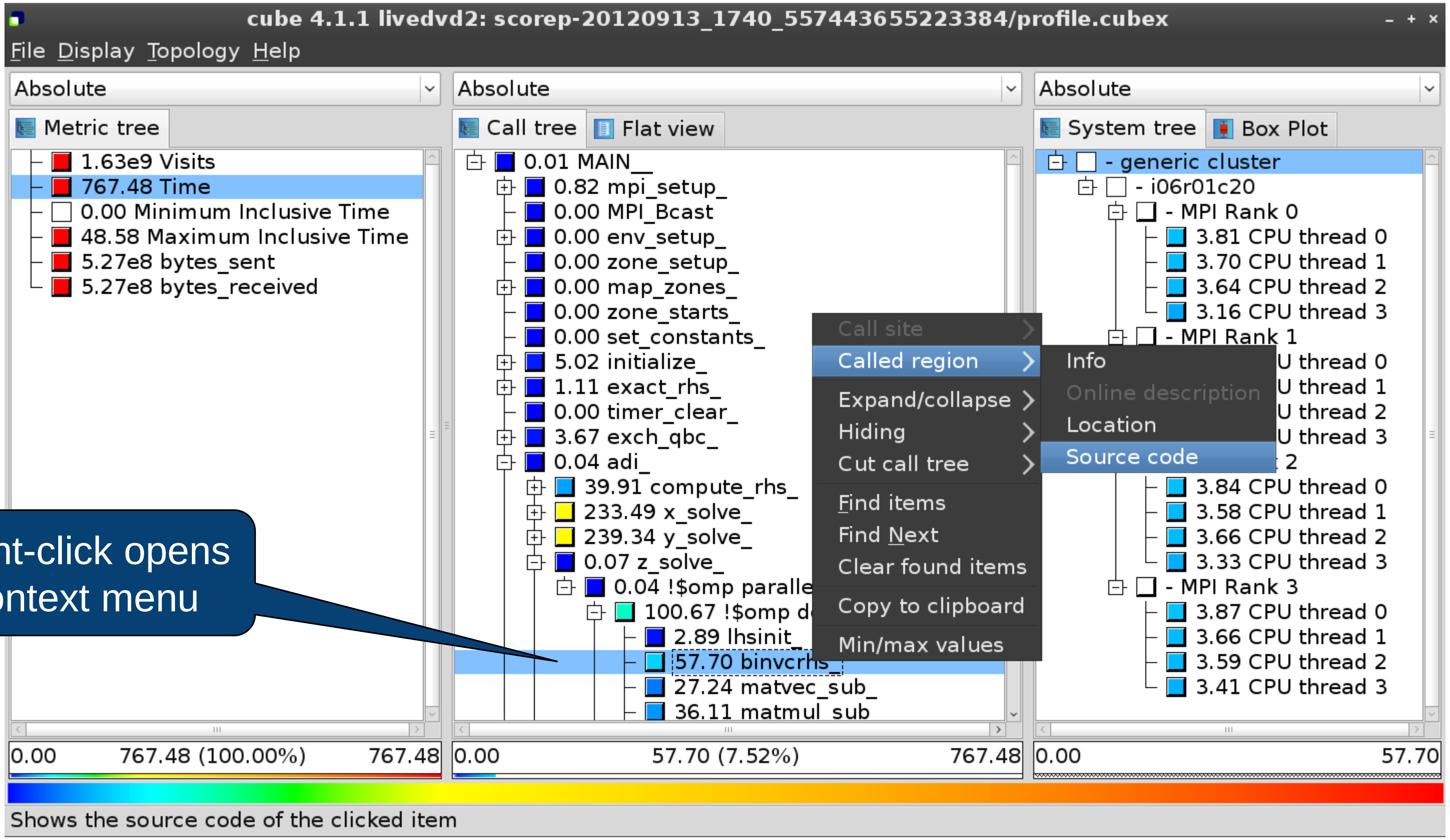
Selecting a call path


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex
File Display Topology Help
Absolute
Absolute
Absolute
Metric tree
Call tree
Flat view
System tree
Box Plot

```
Metric tree:
1.63e9 Visits
767.48 Time
0.00 Minimum Inclusive Time
48.58 Maximum Inclusive Time
5.27e8 bytes_sent
5.27e8 bytes_received

Call tree:
0.01 MAIN__
  0.82 mpi_setup_
  0.00 MPI_Bcast
  0.00 env_setup_
  0.00 zone_setup_
  0.00 map_zones_
  0.00 zone_starts_
  0.00 set_constants_
  5.02 initialize_
  1.11 exact_rhs_
  0.00 timer_clear_
  3.67 exch_qbc_
  0.04 adi_
    39.91 compute_rhs_
    233.49 x_solve_
    239.34 y_solve_
    0.07 z_solve_
      0.04 !$omp parallel @z_solve.f:43
        100.67 !$omp do @z_solve.f:52
          2.89 lhsinit_
          57.70 binvcrhs_
          27.24 matvec_sub_
          36.11 matmul_sub_

System tree:
- generic cluster
  - i06r01c20
    - MPI Rank 0
      3.81 CPU thread 0
      3.70 CPU thread 1
      3.64 CPU thread 2
      3.16 CPU thread 3
    - MPI Rank 1
      3.83 CPU thread 0
      3.29 CPU thread 1
      3.72 CPU thread 2
      3.62 CPU thread 3
    - MPI Rank 2
      3.84 CPU thread 0
      3.58 CPU thread 1
      3.66 CPU thread 2
      3.33 CPU thread 3
    - MPI Rank 3
      3.87 CPU thread 0
      3.66 CPU thread 1
      3.59 CPU thread 2
      3.41 CPU thread 3

0.00   767.48 (100.00%)   767.48
0.00   57.70 (7.52%)   767.48
0.00   57.70
```

Selection updates metric values shown in columns to the right

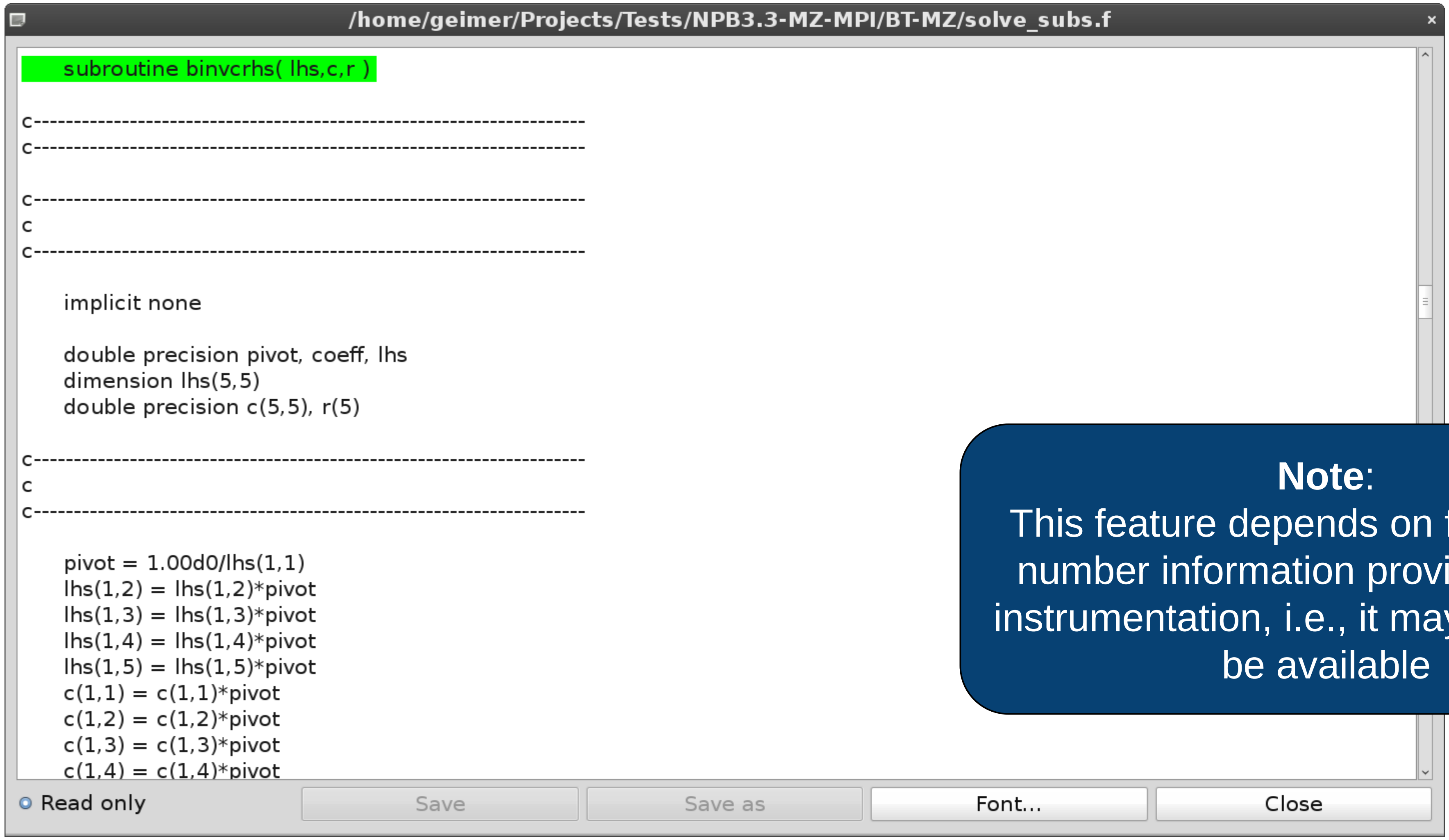
Source-code view via context menu


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

Shows the source code of the clicked item

Right-click opens context menu

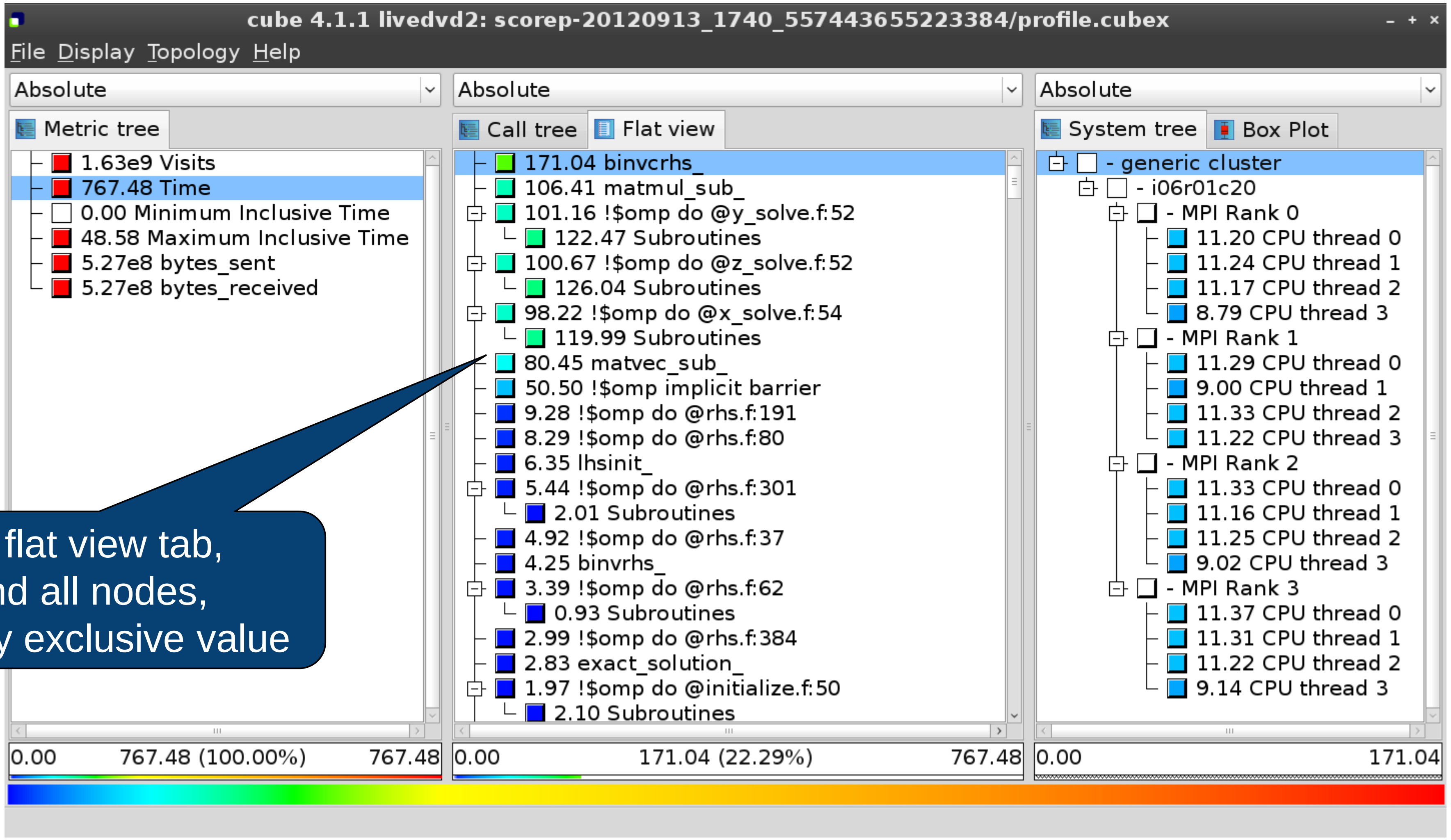
Source-code view


/home/geimer/Projects/Tests/NPB3.3-MZ-MPI/BT-MZ/solve_subs.f
Read only
Save
Save as
Font...
Close

```
      subroutine binvcrhs( lhs,c,r )

c---------------------------------------------------------------------
c---------------------------------------------------------------------

c---------------------------------------------------------------------
c
c---------------------------------------------------------------------

      implicit none

      double precision pivot, coeff, lhs
      dimension lhs(5,5)
      double precision c(5,5), r(5)

c---------------------------------------------------------------------
c
c---------------------------------------------------------------------

      pivot = 1.00d0/lhs(1,1)
      lhs(1,2) = lhs(1,2)*pivot
      lhs(1,3) = lhs(1,3)*pivot
      lhs(1,4) = lhs(1,4)*pivot
      lhs(1,5) = lhs(1,5)*pivot
      c(1,1) = c(1,1)*pivot
      c(1,2) = c(1,2)*pivot
      c(1,3) = c(1,3)*pivot
      c(1,4) = c(1,4)*pivot
```

Note:
This feature depends on file and line number information provided by the instrumentation, i.e., it may not always be available

Flat profile view


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

File Display Topology Help

Absolute | Absolute | Absolute

Metric tree

```
1.63e9 Visits
767.48 Time
0.00 Minimum Inclusive Time
48.58 Maximum Inclusive Time
5.27e8 bytes_sent
5.27e8 bytes_received
```

Call tree | Flat view

```
171.04 binvcrhs_
106.41 matmul_sub_
101.16 !$omp do @y_solve.f:52
  122.47 Subroutines
100.67 !$omp do @z_solve.f:52
  126.04 Subroutines
98.22 !$omp do @x_solve.f:54
  119.99 Subroutines
80.45 matvec_sub_
50.50 !$omp implicit barrier
9.28 !$omp do @rhs.f:191
8.29 !$omp do @rhs.f:80
6.35 lhsinit_
5.44 !$omp do @rhs.f:301
  2.01 Subroutines
4.92 !$omp do @rhs.f:37
4.25 binvrhs_
3.39 !$omp do @rhs.f:62
  0.93 Subroutines
2.99 !$omp do @rhs.f:384
2.83 exact_solution_
1.97 !$omp do @initialize.f:50
  2.10 Subroutines
```

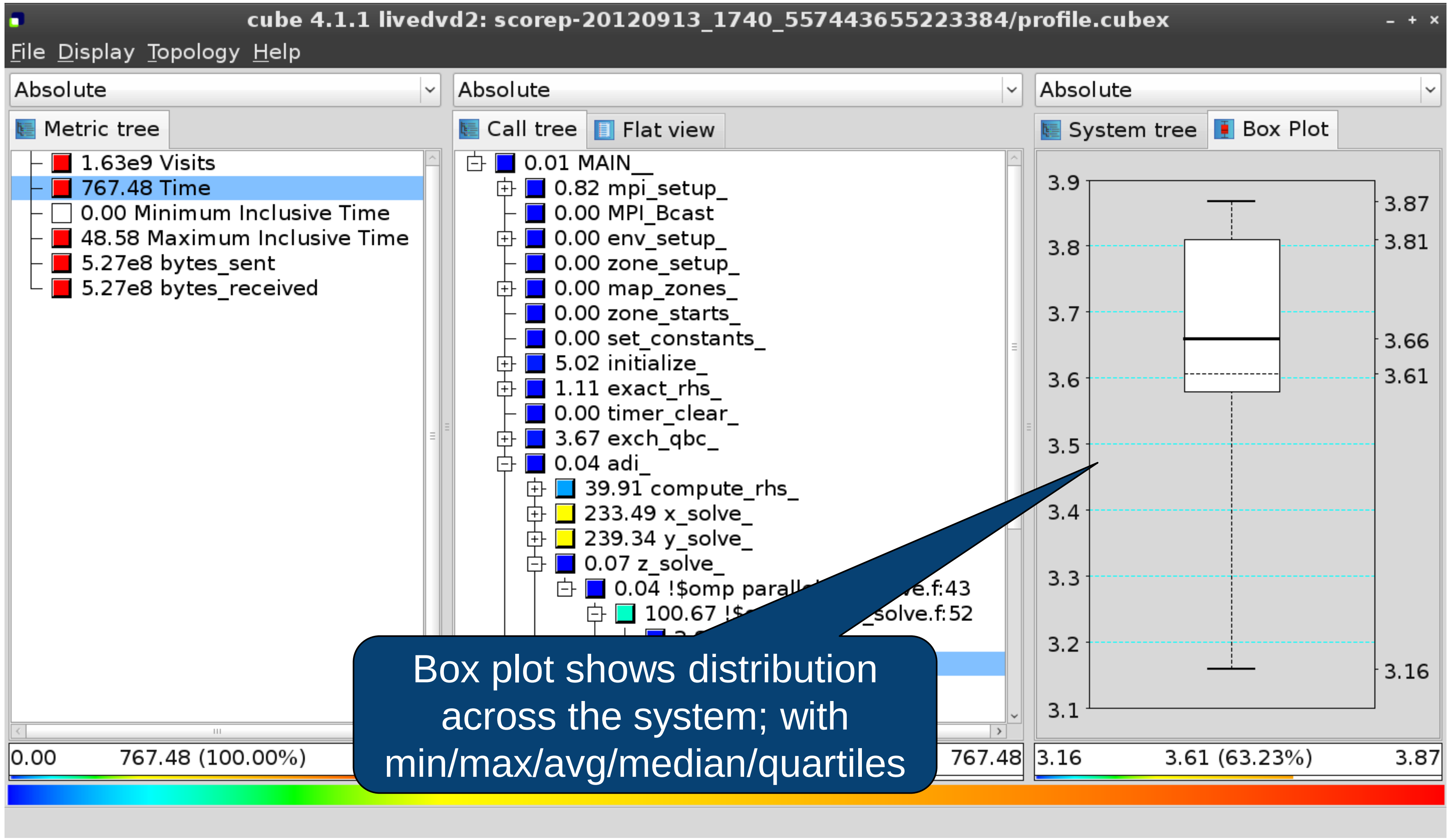
System tree | Box Plot

```
- generic cluster
  - i06r01c20
    - MPI Rank 0
      11.20 CPU thread 0
      11.24 CPU thread 1
      11.17 CPU thread 2
      8.79 CPU thread 3
    - MPI Rank 1
      11.29 CPU thread 0
      9.00 CPU thread 1
      11.33 CPU thread 2
      11.22 CPU thread 3
    - MPI Rank 2
      11.33 CPU thread 0
      11.16 CPU thread 1
      11.25 CPU thread 2
      9.02 CPU thread 3
    - MPI Rank 3
      11.37 CPU thread 0
      11.31 CPU thread 1
      11.22 CPU thread 2
      9.14 CPU thread 3
```

```
0.00    767.48 (100.00%)    767.48
0.00    171.04 (22.29%)    767.48
0.00    171.04
```

Select flat view tab,
expand all nodes,
and sort by exclusive value

Box plot view


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

File Display Topology Help

Absolute | Absolute | Absolute

Metric tree | Call tree | Flat view | System tree | Box Plot

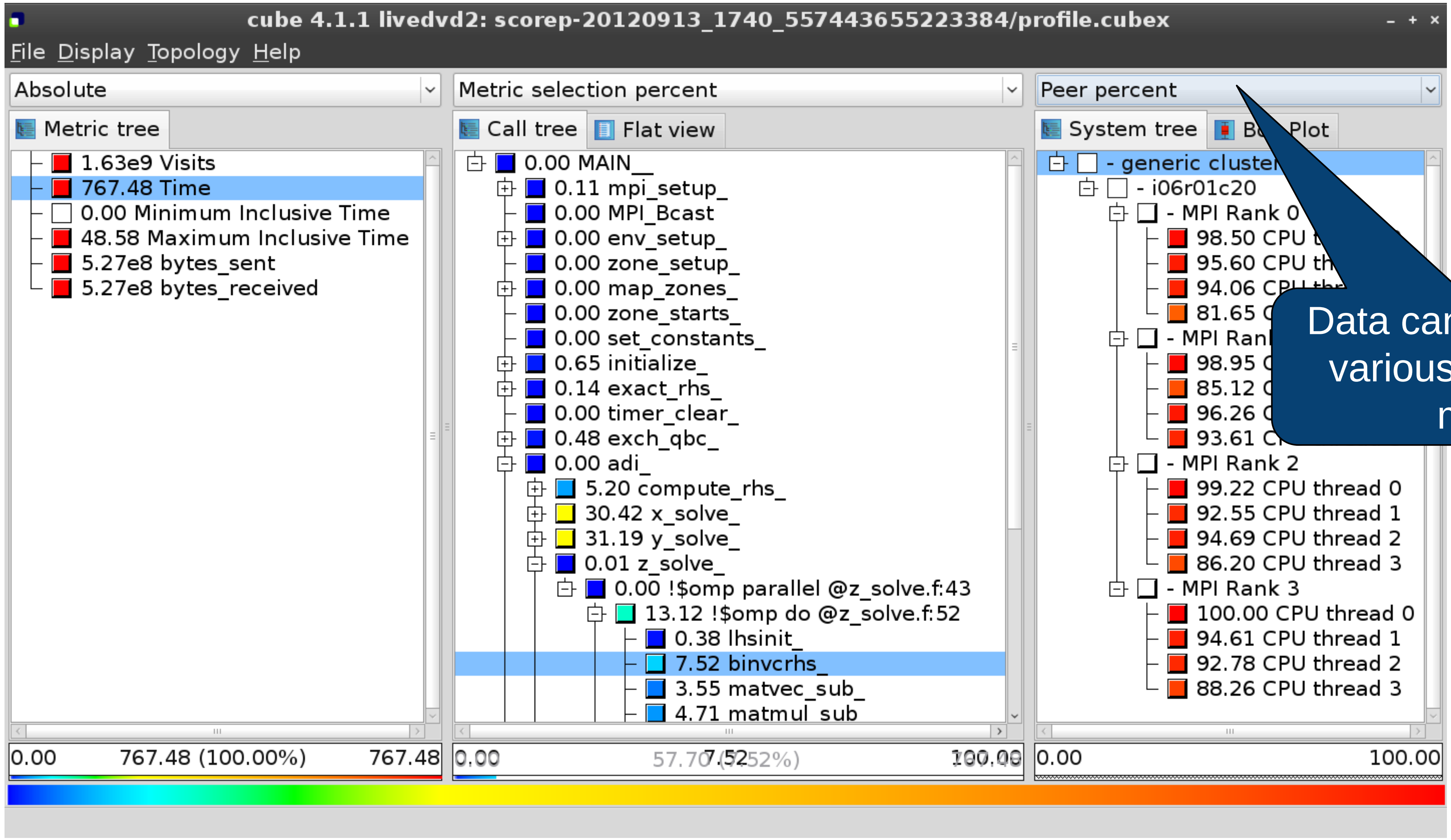
```
1.63e9 Visits
767.48 Time
0.00 Minimum Inclusive Time
48.58 Maximum Inclusive Time
5.27e8 bytes_sent
5.27e8 bytes_received
```

```
0.01 MAIN__
  0.82 mpi_setup_
  0.00 MPI_Bcast
  0.00 env_setup_
  0.00 zone_setup_
  0.00 map_zones_
  0.00 zone_starts_
  0.00 set_constants_
  5.02 initialize_
  1.11 exact_rhs_
  0.00 timer_clear_
  3.67 exch_qbc_
  0.04 adi_
    39.91 compute_rhs_
    233.49 x_solve_
    239.34 y_solve_
    0.07 z_solve_
      0.04 !$omp parall        e.f:43
        100.67 !$              _solve.f:52
```

```
0.00        767.48 (100.00%)        767.48   3.16      3.61 (63.23%)      3.87
```

Box plot shows distribution across the system; with min/max/avg/median/quartiles

Alternative display modes



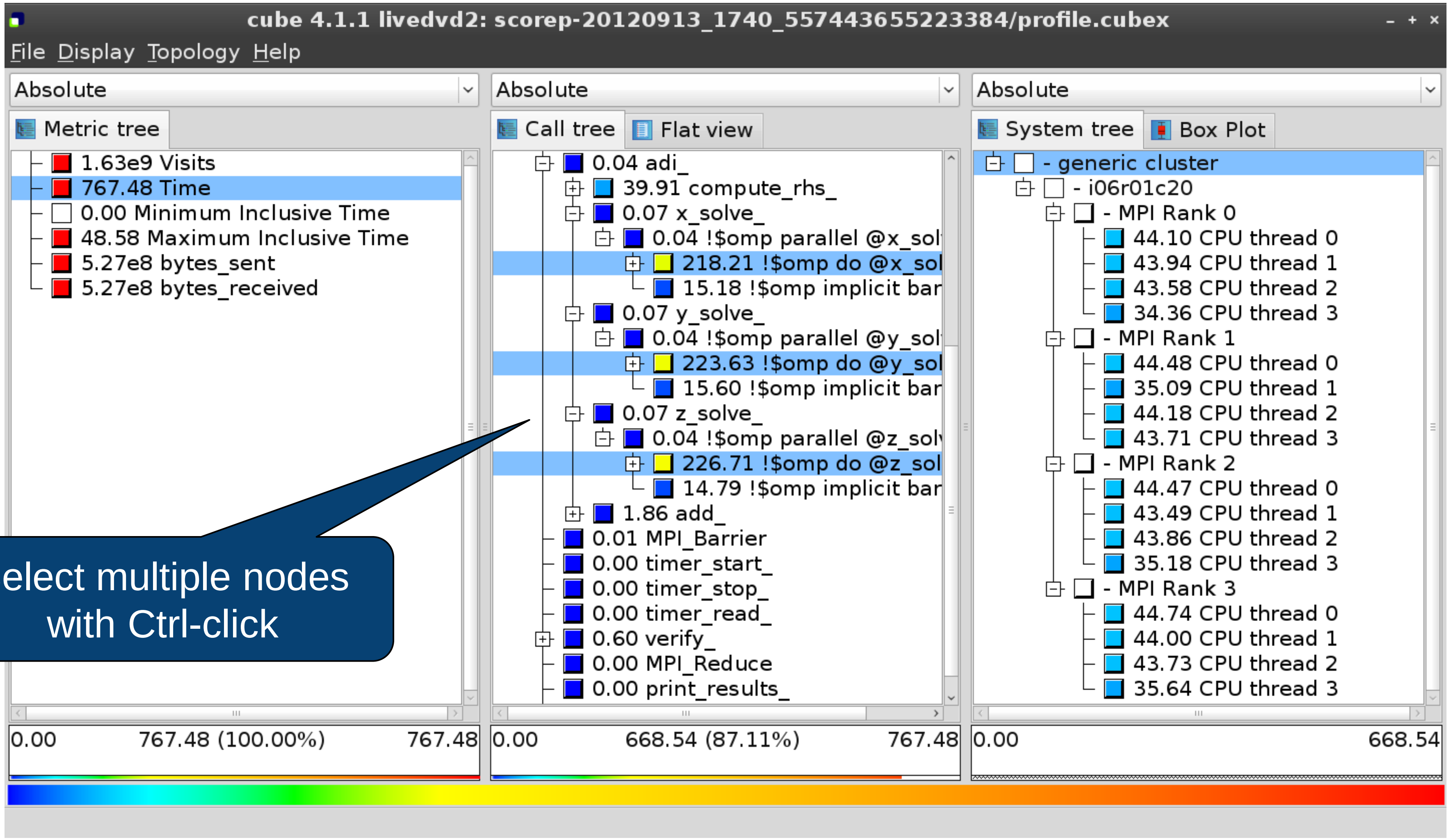
```
cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex
File  Display  Topology  Help

Absolute                         Metric selection percent          Peer percent
Metric tree                      Call tree | Flat view             System tree | Box Plot

1.63e9 Visits                    0.00 MAIN__                        - generic cluster
767.48 Time                        0.11 mpi_setup_                    - i06r01c20
0.00 Minimum Inclusive Time        0.00 MPI_Bcast                       - MPI Rank 0
48.58 Maximum Inclusive Time       0.00 env_setup_                        98.50 CPU t
5.27e8 bytes_sent                  0.00 zone_setup_                       95.60 CPU th
5.27e8 bytes_received              0.00 map_zones_                        94.06 CPU thr
                                   0.00 zone_starts_                      81.65 C
                                   0.00 set_constants_                  - MPI Rank
                                   0.65 initialize_                       98.95 C
                                   0.14 exact_rhs_                        85.12 C
                                   0.00 timer_clear_                      96.26 C
                                   0.48 exch_qbc_                         93.61 C
                                   0.00 adi_                            - MPI Rank 2
                                     5.20 compute_rhs_                    99.22 CPU thread 0
                                     30.42 x_solve_                       92.55 CPU thread 1
                                     31.19 y_solve_                       94.69 CPU thread 2
                                     0.01 z_solve_                        86.20 CPU thread 3
                                       0.00 !$omp parallel @z_solve.f:43 - MPI Rank 3
                                         13.12 !$omp do @z_solve.f:52     100.00 CPU thread 0
                                           0.38 lhsinit_                  94.61 CPU thread 1
                                           7.52 binvcrhs_                 92.78 CPU thread 2
                                           3.55 matvec_sub_               88.26 CPU thread 3
                                           4.71 matmul_sub

0.00   767.48 (100.00%)   767.48   0.00   7.52   100.00           0.00   100.00
```

Data can be shown in various percentage modes

Important display modes

- Absolute
  - Absolute value shown in seconds/bytes/counts
- Selection percent
  - Value shown as percentage w.r.t. the selected node "on the left" (metric/call path)
- Peer percent (system tree only)
  - Value shown as percentage relative to the maximum peer value

Multiple selection

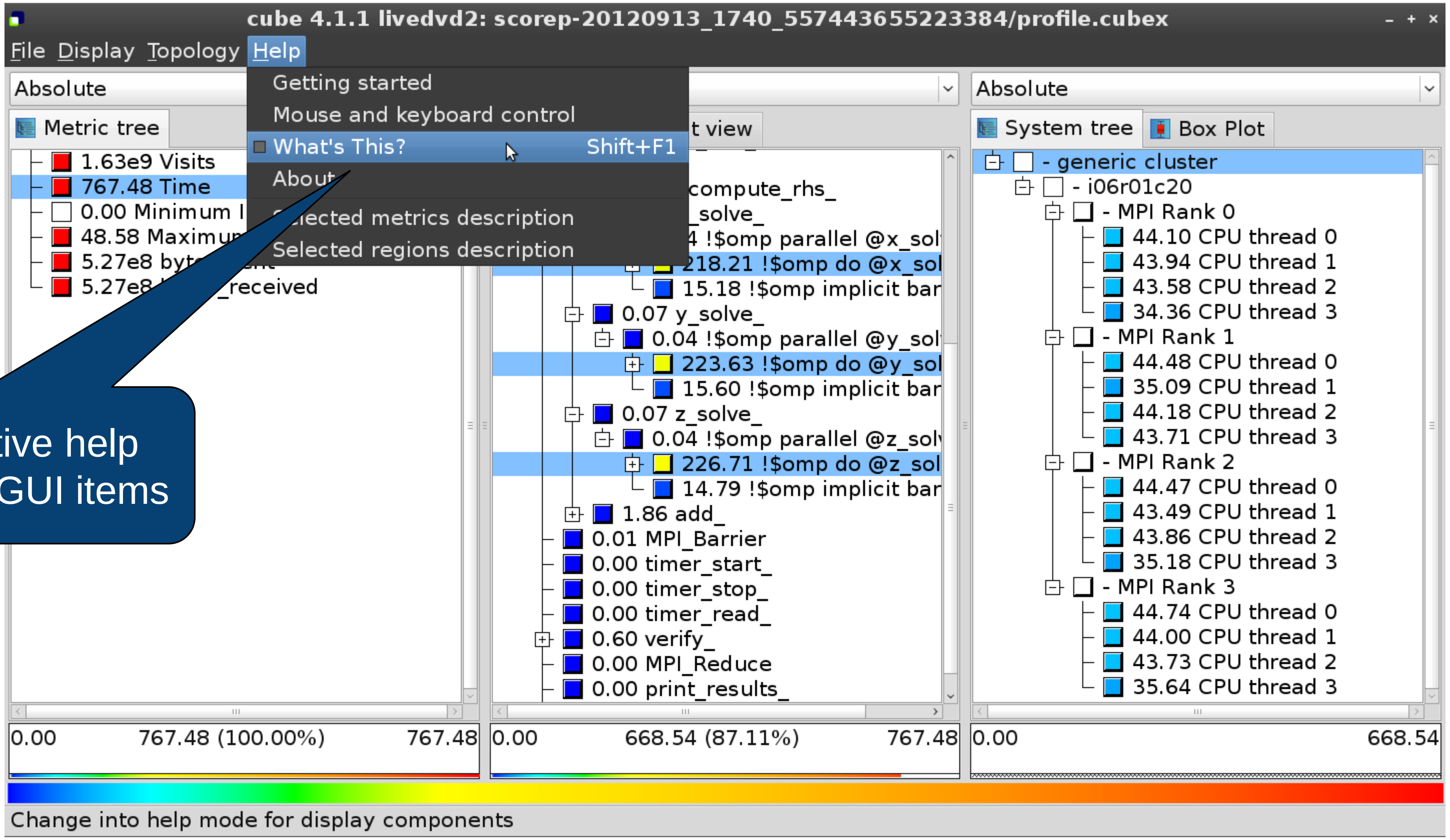

cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex

File Display Topology Help

```
Absolute                          Absolute                                Absolute
Metric tree                       Call tree | Flat view                   System tree | Box Plot

1.63e9 Visits                     0.04 adi_                               - generic cluster
767.48 Time                         39.91 compute_rhs_                      - i06r01c20
0.00 Minimum Inclusive Time         0.07 x_solve_                             - MPI Rank 0
48.58 Maximum Inclusive Time          0.04 !$omp parallel @x_sol                44.10 CPU thread 0
5.27e8 bytes_sent                       218.21 !$omp do @x_sol                  43.94 CPU thread 1
5.27e8 bytes_received                   15.18 !$omp implicit bar                43.58 CPU thread 2
                                    0.07 y_solve_                               34.36 CPU thread 3
                                      0.04 !$omp parallel @y_sol              - MPI Rank 1
                                        223.63 !$omp do @y_sol                  44.48 CPU thread 0
                                        15.60 !$omp implicit bar                35.09 CPU thread 1
                                    0.07 z_solve_                               44.18 CPU thread 2
                                      0.04 !$omp parallel @z_sol                43.71 CPU thread 3
                                        226.71 !$omp do @z_sol                - MPI Rank 2
                                        14.79 !$omp implicit bar                44.47 CPU thread 0
                                    1.86 add_                                   43.49 CPU thread 1
                                  0.01 MPI_Barrier                              43.86 CPU thread 2
                                  0.00 timer_start_                             35.18 CPU thread 3
                                  0.00 timer_stop_                            - MPI Rank 3
                                  0.00 timer_read_                              44.74 CPU thread 0
                                  0.60 verify_                                  44.00 CPU thread 1
                                  0.00 MPI_Reduce                               43.73 CPU thread 2
                                  0.00 print_results_                           35.64 CPU thread 3

0.00   767.48 (100.00%)   767.48  0.00   668.54 (87.11%)   767.48     0.00                 668.54
```

Select multiple nodes with Ctrl-click

Context-sensitive help


cube 4.1.1 livedvd2: scorep-20120913_1740_557443655223384/profile.cubex
File Display Topology Help
Getting started
Mouse and keyboard control
What's This? Shift+F1
About
Selected metrics description
Selected regions description
Absolute
Metric tree
System tree
Box Plot

```
1.63e9 Visits
767.48 Time
0.00 Minimum I
48.58 Maximum
5.27e8 byt
5.27e8 _received

compute_rhs_
solve_
4 !$omp parallel @x_sol
218.21 !$omp do @x_sol
15.18 !$omp implicit bar
0.07 y_solve_
0.04 !$omp parallel @y_sol
223.63 !$omp do @y_sol
15.60 !$omp implicit bar
0.07 z_solve_
0.04 !$omp parallel @z_sol
226.71 !$omp do @z_sol
14.79 !$omp implicit bar
1.86 add_
0.01 MPI_Barrier
0.00 timer_start_
0.00 timer_stop_
0.00 timer_read_
0.60 verify_
0.00 MPI_Reduce
0.00 print_results_

- generic cluster
  - i06r01c20
    - MPI Rank 0
      44.10 CPU thread 0
      43.94 CPU thread 1
      43.58 CPU thread 2
      34.36 CPU thread 3
    - MPI Rank 1
      44.48 CPU thread 0
      35.09 CPU thread 1
      44.18 CPU thread 2
      43.71 CPU thread 3
    - MPI Rank 2
      44.47 CPU thread 0
      43.49 CPU thread 1
      43.86 CPU thread 2
      35.18 CPU thread 3
    - MPI Rank 3
      44.74 CPU thread 0
      44.00 CPU thread 1
      43.73 CPU thread 2
      35.64 CPU thread 3

0.00   767.48 (100.00%)   767.48
0.00   668.54 (87.11%)    767.48
0.00                      668.54

Change into help mode for display components
```

Context-sensitive help available for all GUI items

Derived metrics

- Derived metrics are defined using CubePL expressions, e.g.:

  metric::time(i)/metric::visits(e)
- Values of derived metrics are not stored, but calculated on-the-fly
- Types of derived metrics:
  - Prederived: evaluation of the CubePL expression is performed before aggregation
  - Postderived: evaluation of the CubePL expression is performed after aggregation

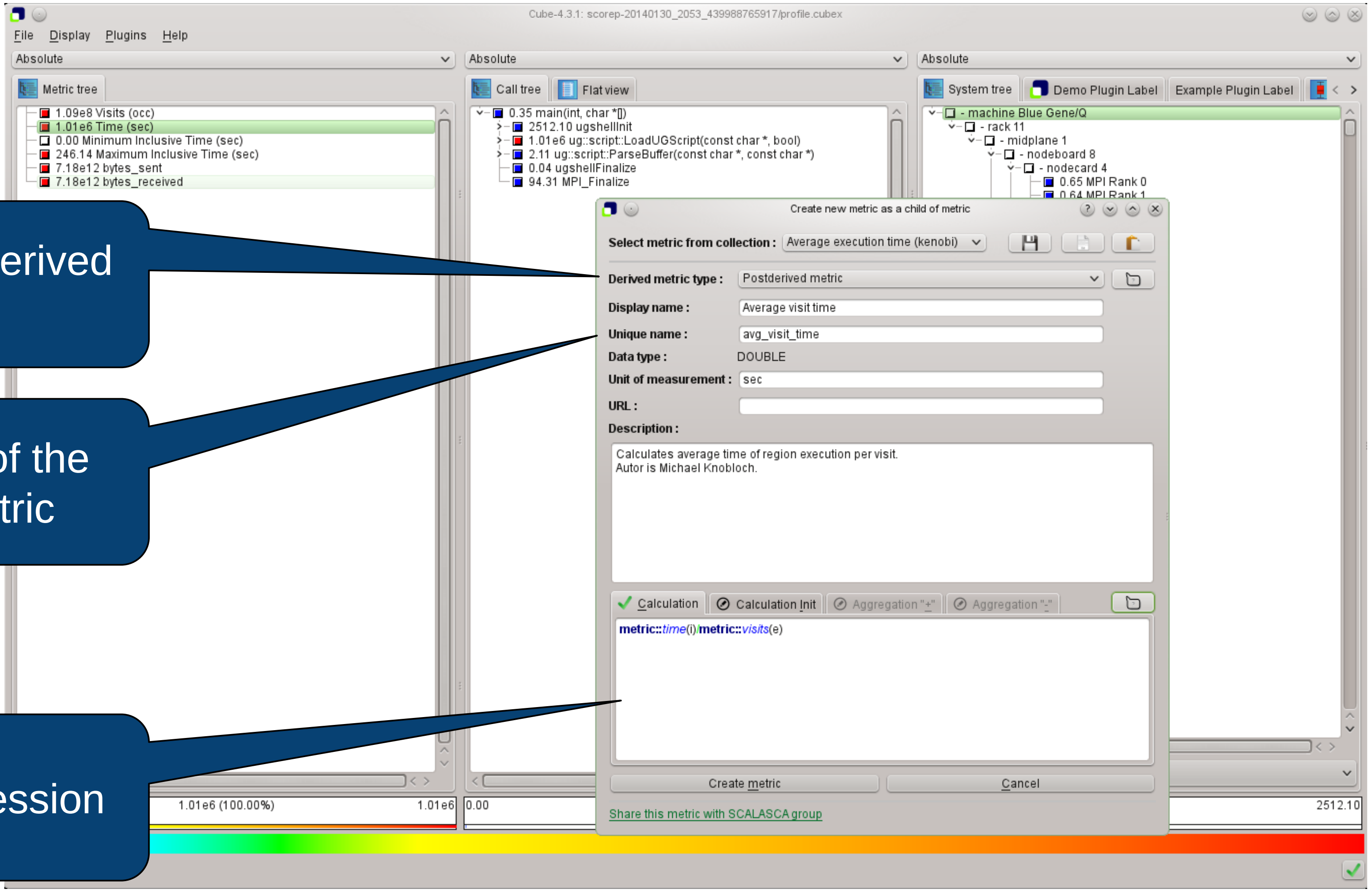
- Examples:
  - "Average execution time": Postderived metric with expression

    metric::time(i)/metric::visits(e)

  - "Number of FLOP per second": Postderived metric with expression

    metric::FLOP()/metric::time()

Derived metrics in Cube GUI



1.09e8 Visits (occ)
1.01e6 Time (sec)
0.00 Minimum Inclusive Time (sec)
246.14 Maximum Inclusive Time (sec)
7.18e12 bytes_sent
7.18e12 bytes_received

0.35 main(int, char *[])
2512.10 ugshellInit
1.01e6 ug::script::LoadUGScript(const char *, bool)
2.11 ug::script::ParseBuffer(const char *, const char *)
0.04 ugshellFinalize
94.31 MPI_Finalize

- machine Blue Gene/Q
- rack 11
- midplane 1
- nodeboard 8
- nodecard 4
0.65 MPI Rank 0
0.64 MPI Rank 1

Select metric from collection : Average execution time (kenobi)
Derived metric type : Postderived metric
Display name : Average visit time
Unique name : avg_visit_time
Data type : DOUBLE
Unit of measurement : sec
URL :
Description :
Calculates average time of region execution per visit.
Autor is Michael Knobloch.

```
metric::time(i)/metric::visits(e)
```

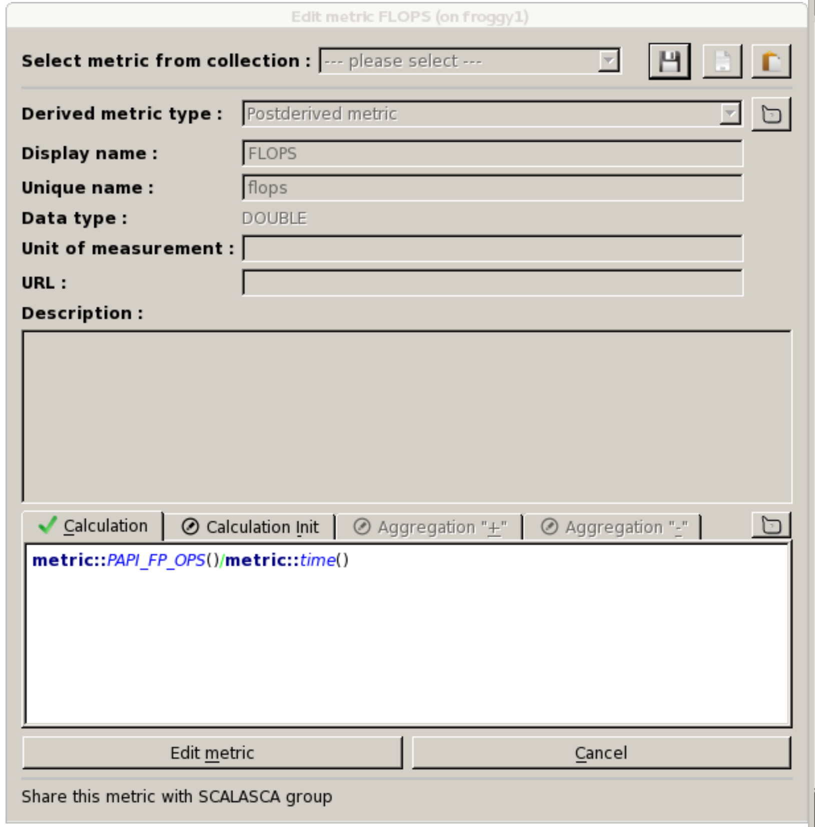
1.01e6 (100.00%) 1.01e6 0.00 2512.10

Collection of derived metrics

Parameters of the derived metric

CubePL expression

Example: FLOPS based on PAPI_FP_OPS and time



Edit metric FLOPS (on froggy1)

Select metric from collection : --- please select ---

Derived metric type : Postderived metric

Display name : FLOPS

Unique name : flops

Data type : DOUBLE

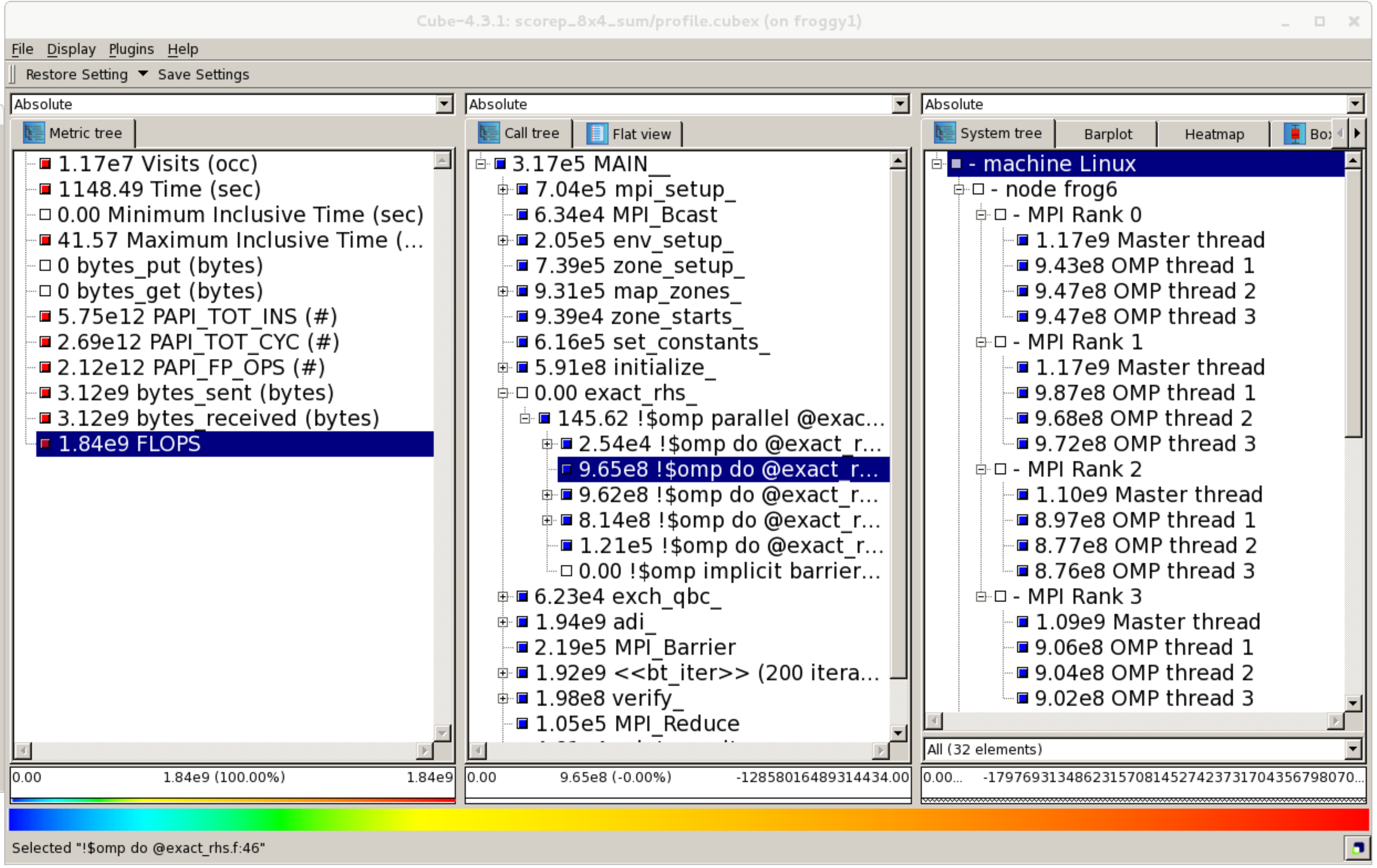
Unit of measurement :

URL :

Description :

```
metric::PAPI_FP_OPS()/metric::time()
```

Share this metric with SCALASCA group



Cube-4.3.1: scorep_8x4_sum/profile.cubex (on froggy1)

```
1.17e7 Visits (occ)
1148.49 Time (sec)
0.00 Minimum Inclusive Time (sec)
41.57 Maximum Inclusive Time (...
0 bytes_put (bytes)
0 bytes_get (bytes)
5.75e12 PAPI_TOT_INS (#)
2.69e12 PAPI_TOT_CYC (#)
2.12e12 PAPI_FP_OPS (#)
3.12e9 bytes_sent (bytes)
3.12e9 bytes_received (bytes)
1.84e9 FLOPS
```

```
3.17e5 MAIN__
  7.04e5 mpi_setup_
  6.34e4 MPI_Bcast
  2.05e5 env_setup_
  7.39e5 zone_setup_
  9.31e5 map_zones_
  9.39e4 zone_starts_
  6.16e5 set_constants_
  5.91e8 initialize_
  0.00 exact_rhs_
    145.62 !$omp parallel @exac...
      2.54e4 !$omp do @exact_r...
      9.65e8 !$omp do @exact_r...
      9.62e8 !$omp do @exact_r...
      8.14e8 !$omp do @exact_r...
      1.21e5 !$omp do @exact_r...
      0.00 !$omp implicit barrier...
  6.23e4 exch_qbc_
  1.94e9 adi_
  2.19e5 MPI_Barrier
  1.92e9 <<bt_iter>> (200 itera...
  1.98e8 verify_
  1.05e5 MPI_Reduce
```

```
- machine Linux
  - node frog6
    - MPI Rank 0
      1.17e9 Master thread
      9.43e8 OMP thread 1
      9.47e8 OMP thread 2
      9.47e8 OMP thread 3
    - MPI Rank 1
      1.17e9 Master thread
      9.87e8 OMP thread 1
      9.68e8 OMP thread 2
      9.72e8 OMP thread 3
    - MPI Rank 2
      1.10e9 Master thread
      8.97e8 OMP thread 1
      8.77e8 OMP thread 2
      8.76e8 OMP thread 3
    - MPI Rank 3
      1.09e9 Master thread
      9.06e8 OMP thread 1
      9.04e8 OMP thread 2
      9.02e8 OMP thread 3
```

All (32 elements)

0.00 1.84e9 (100.00%) 1.84e9

0.00 9.65e8 (-0.00%) -12858016489314434.00

0.00... -179769313486231570814527423731704356798070...

Selected "!$omp do @exact_rhs.f:46"

Iteration profiling

- Show time dependent behavior by “unrolling” iterations

- Preparations:
  - Mark loop body by using Score-P instrumentation API in your source code

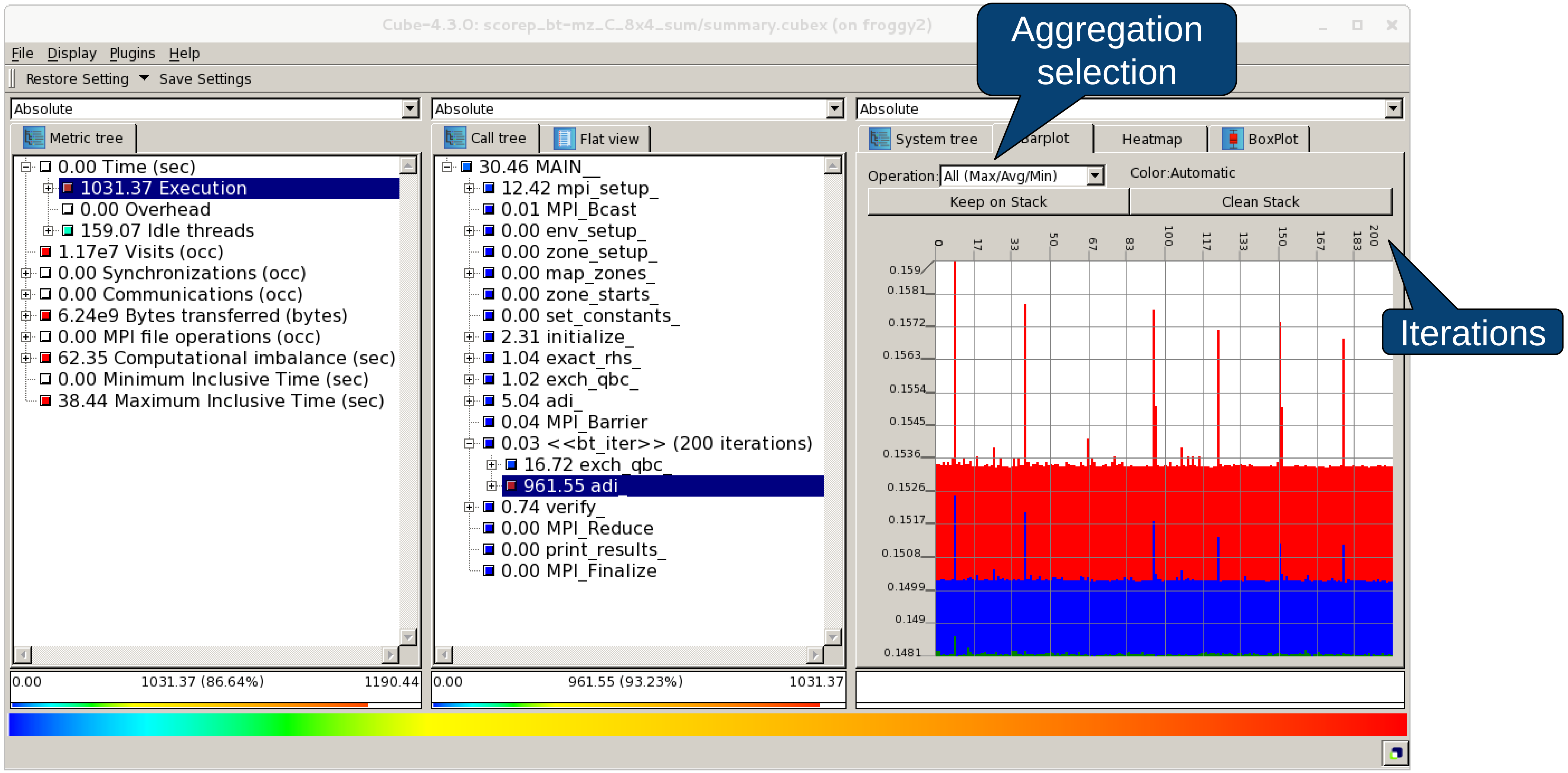
```
SCOREP_USER_REGION_DEFINE( scorep_bt_loop )
SCOREP_USER_REGION_BEGIN( scorep_bt_loop, "<<bt_iter>>", SCOREP_USER_REGION_TYPE_DYNAMIC )
SCOREP_USER_REGION_END( scorep_bt_loop )
```

- Result in the Cube profile:
  - Iterations shown as separate call trees
  - ➢ Useful for checking results for specific iterations

or

  - Select your user-instrumented region and mark it as loop
  - Choose “Hide iterations”
  - ➢ View the Barplot statistics or the (thread x iterations) Heatmap

Iteration profiling: Barplot



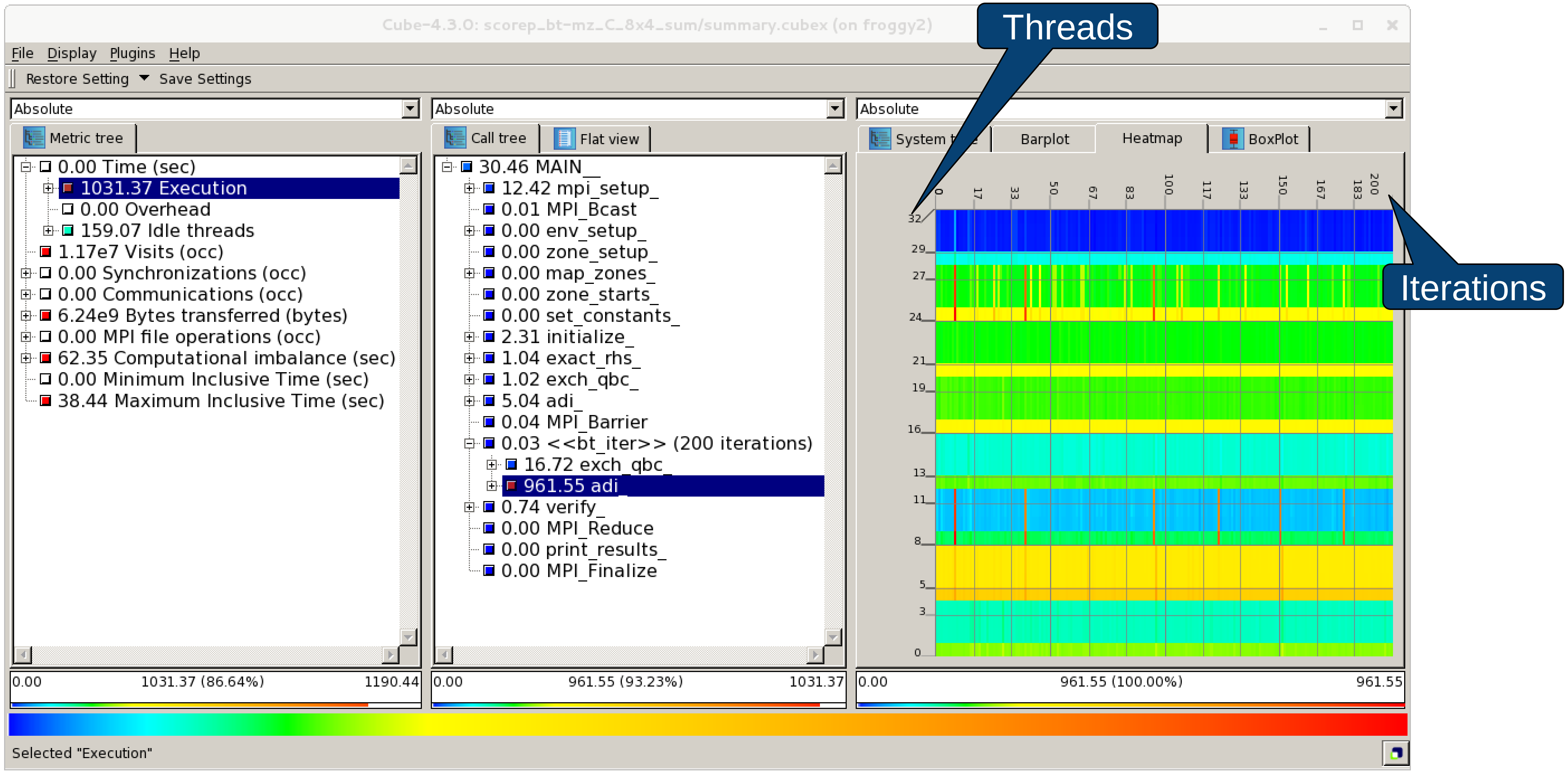
Metric tree:
- 0.00 Time (sec)
  - 1031.37 Execution
  - 0.00 Overhead
  - 159.07 Idle threads
- 1.17e7 Visits (occ)
- 0.00 Synchronizations (occ)
- 0.00 Communications (occ)
- 6.24e9 Bytes transferred (bytes)
- 0.00 MPI file operations (occ)
- 62.35 Computational imbalance (sec)
- 0.00 Minimum Inclusive Time (sec)
- 38.44 Maximum Inclusive Time (sec)

Call tree:
- 30.46 MAIN_
  - 12.42 mpi_setup_
  - 0.01 MPI_Bcast
  - 0.00 env_setup_
  - 0.00 zone_setup_
  - 0.00 map_zones_
  - 0.00 zone_starts_
  - 0.00 set_constants_
  - 2.31 initialize_
  - 1.04 exact_rhs_
  - 1.02 exch_qbc_
  - 5.04 adi_
  - 0.04 MPI_Barrier
  - 0.03 <<bt_iter>> (200 iterations)
    - 16.72 exch_qbc_
    - 961.55 adi_
  - 0.74 verify_
  - 0.00 MPI_Reduce
  - 0.00 print_results_
  - 0.00 MPI_Finalize

0.00 1031.37 (86.64%) 1190.44

0.00 961.55 (93.23%) 1031.37

Iteration profiling: Heatmap



Cube-4.3.0: scorep_bt-mz_C_8x4_sum/summary.cubex (on froggy2)

```
0.00 Time (sec)
  1031.37 Execution
  0.00 Overhead
  159.07 Idle threads
1.17e7 Visits (occ)
0.00 Synchronizations (occ)
0.00 Communications (occ)
6.24e9 Bytes transferred (bytes)
0.00 MPI file operations (occ)
62.35 Computational imbalance (sec)
0.00 Minimum Inclusive Time (sec)
38.44 Maximum Inclusive Time (sec)
```

```
30.46 MAIN__
  12.42 mpi_setup_
  0.01 MPI_Bcast
  0.00 env_setup_
  0.00 zone_setup_
  0.00 map_zones_
  0.00 zone_starts_
  0.00 set_constants_
  2.31 initialize_
  1.04 exact_rhs_
  1.02 exch_qbc_
  5.04 adi_
  0.04 MPI_Barrier
  0.03 <<bt_iter>> (200 iterations)
    16.72 exch_qbc_
    961.55 adi_
  0.74 verify_
  0.00 MPI_Reduce
  0.00 print_results_
  0.00 MPI_Finalize
```

0.00 1031.37 (86.64%) 1190.44

0.00 961.55 (93.23%) 1031.37

0.00 961.55 (100.00%) 961.55

Selected "Execution"

CUBE algebra utilities

- Extracting solver sub-tree from analysis report

```
% cube_cut  -r '<<ITERATION>>'  scorep_bt-mz_C_16x8_sum/profile.cubex
Writing cut.cubex... done.
```

- Calculating difference of two reports

```
% cube_diff  scorep_bt-mz_C_16x8_sum/profile.cubex  cut.cubex
Writing diff.cubex... done.
```

- Additional utilities for merging, calculating mean, etc.
- Default output of cube_*utility* is a new report *utility*.cubex
- Further utilities for report scoring & statistics
- Run utility with `` `-h’ `` (or no arguments) for brief usage info

Square sneak preview

- Scalasca provides ***square*** to facilitate analysis report exploration
  - square = scalasca –examine [OPTIONS] (./scorep_expt_sum | ./profile.cubex)
- Processes intermediate .cubex files produced by Score-P and Scout
  - profile.cubex -> summary.cubex
  - scout.cubex -> trace.cubex
- and (optionally) starts CUBE GUI with the post-processed file
  - containing additional derived metrics and metric hierarchies

Cube: Further information

- Parallel program analysis report exploration tools
  - Libraries for Cube report reading & writing
  - Algebra utilities for report processing
  - GUI for interactive analysis exploration
- Available under 3-clause BSD open-source license
- Documentation & sources:
  - http://www.scalasca.org
- User guide also part of installation:
  - <prefix>/share/doc/CubeGuide.pdf
- Contact:
  - mailto: scalasca@fz-juelich.de

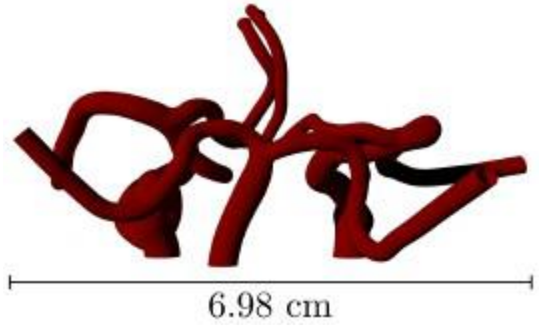
Score-P/CUBE case study HemeLB

HemeLB (SuperMUC-NG: no GPUs)

CompBioMed

DOI 10.5281/zenodo.4105743

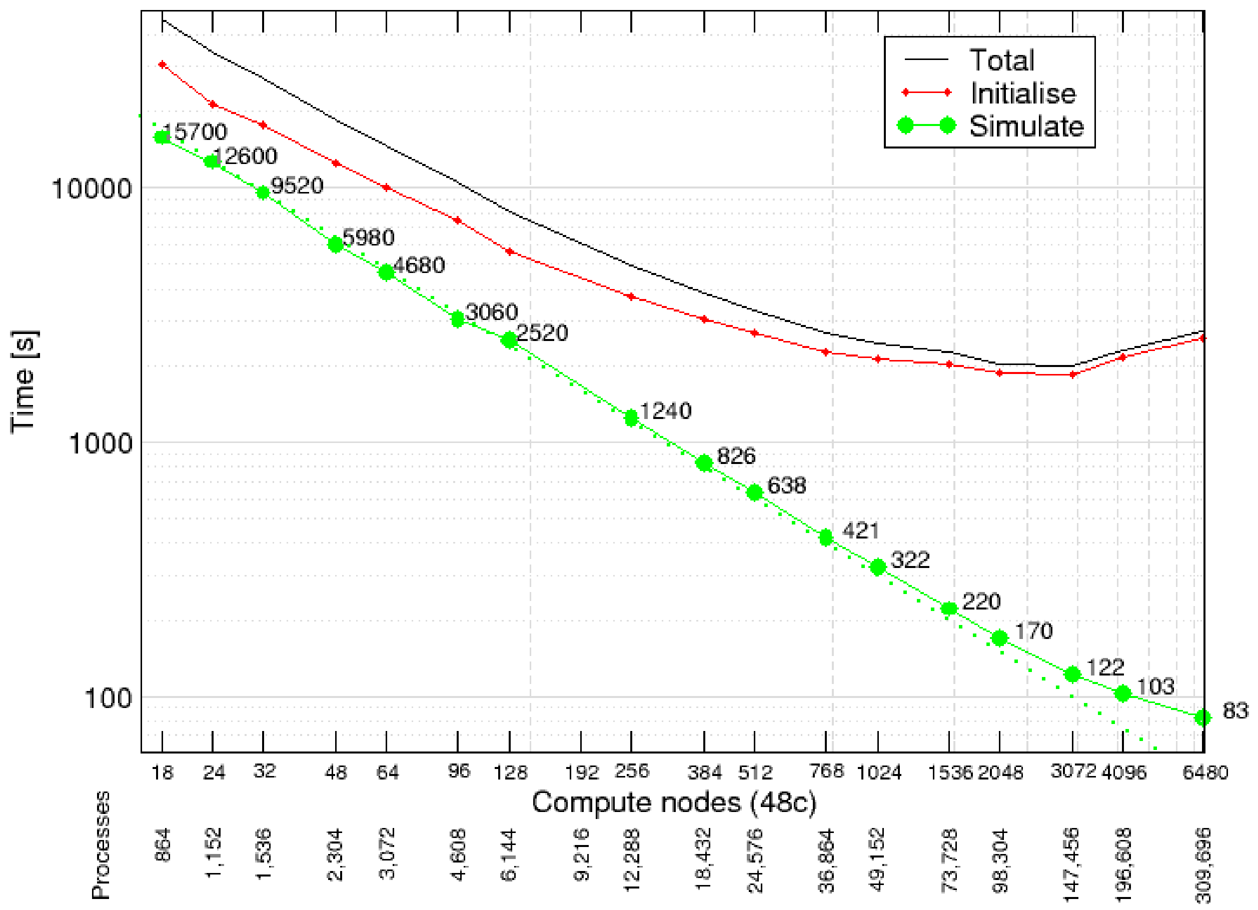
- 3D macroscopic blood flow in human arterial system developed by UC London (UK)
  - lattice-Boltzmann method tracking fluid particles on a lattice grid with complex boundary conditions
  - exascale flagship application of EU H2020 HPC Centre of Excellence for Computational Biomedicine
- HemeLB open-source code and test case: www.hemelb.org
  - C++ parallelized with MPI [+ CUDA unused]
    - Intel Studio 2019u4 compiler and MPI library (v19.0.4.243)
    - configured with 2 'reader' processes (intermediate MPI file writing disabled)
    - MPI-3 shared-memory model employed within compute nodes
      to reduce memory requirements when distributing lattice blocks from reader processes
  - Focus of analysis 5,000 time-step (500μs) simulation of cerebrovascular "circle of Willis" geometry
    - 6.4μm lattice resolution (21.15 GiB): 10,154,448,502 lattice sites

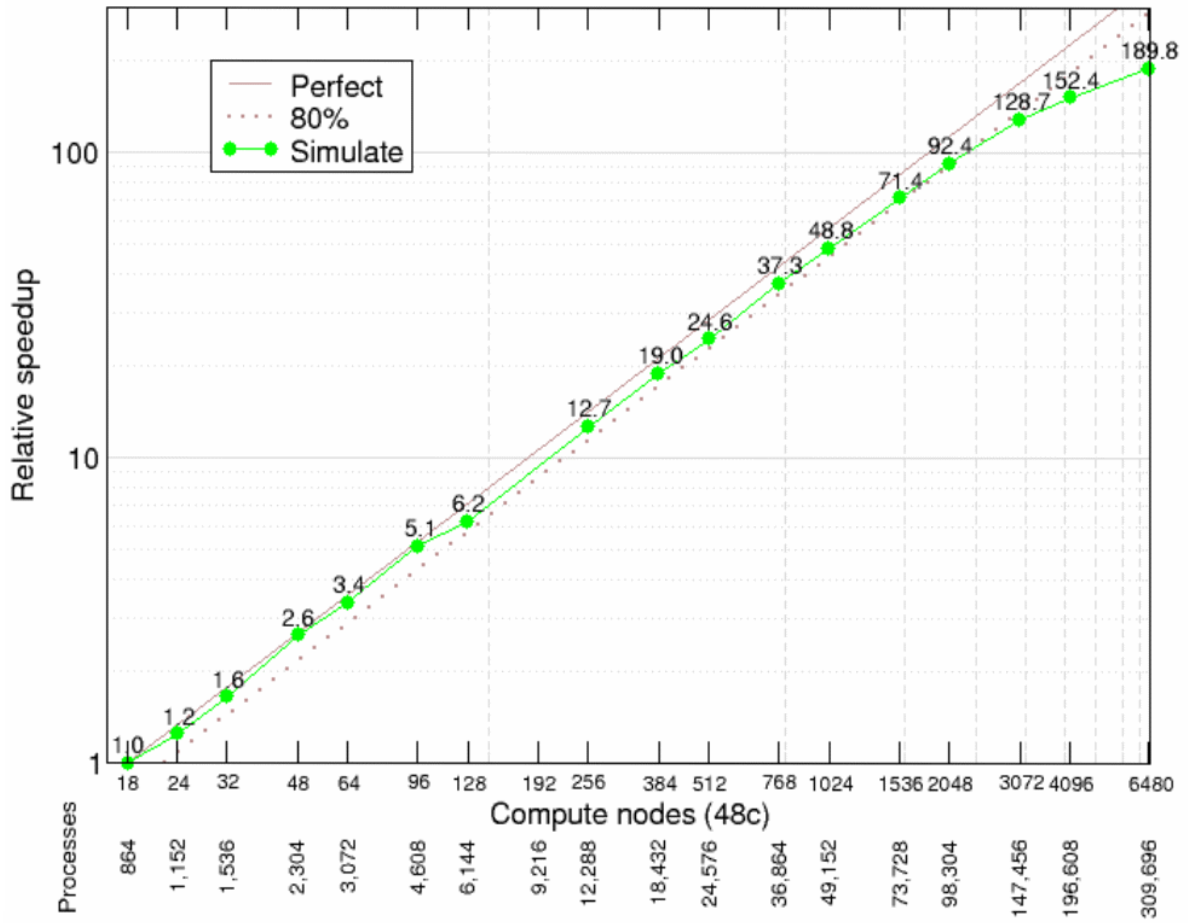


- Executed on *SuperMUC-NG* Lenovo ThinkSystem SD650 (LRZ):
  - 2x 24-core Intel Xeon Platinum 8174 ('Skylake') @ 3.1GHz
  - 48 MPI processes/node, 6452 (of 6480) compute nodes: 309,696 MPI processes
  - 190x speed-up from 864 cores: 80% scaling efficiency to over 100,000 cores

⇒ ***Identification & quantification of impact of load balance and its variation***

HemeLB@SNG strong scaling of FOA *RunSimulation*





[Execution of 9,216 processes on 192 compute nodes not possible due to insufficient compute nodes with adequate memory in 'fat' partition (768 GiB vs. regular 96 GiB node memory]

HemeLB@SNG strong scaling efficiency of FOA *RunSimulation*

Compute nodes	24	32	48	64	96	128	192	256	384	512	768	1024	1536	2048	3072	4096	6452
Processes	**1152**	**1536**	**2304**	**3072**	**4608**	**6144**	**9216**	**12288**	**18432**	**24576**	**36864**	**49152**	**73728**	**98304**	**147456**	**196608**	**309696**
Global scaling efficiency	**0.79**	**0.79**	**0.84**	**0.80**	**0.82**	**0.75**		**0.73**	**0.72**	**0.73**	**0.74**	**0.68**	**0.68**	**0.65**	**0.62**	**0.57**	**0.45**
- Parallel efficiency	0.79	0.80	0.87	0.83	0.86	0.80		0.75	0.74	0.74	0.77	0.71	0.72	0.70	0.72	0.70	0.73
- - Load balance efficiency	0.79	0.80	0.88	0.84	0.86	0.80		0.75	0.74	0.75	0.78	0.72	0.74	0.72	0.74	0.73	0.80
- - Communication efficiency	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	0.99	0.99	0.99	0.98	0.98	0.97	0.96	0.92
- Computation scaling	*1.00*	0.99	0.96	0.96	0.95	0.93		0.98	0.98	0.98	0.96	0.96	0.94	0.93	0.87	0.81	0.61
- - Instructions scaling	*1.00*	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	0.99	0.97	0.94	0.89	0.79	0.67	0.45
- - IPC scaling	*1.00*	0.99	0.96	0.96	0.95	0.93		0.98	0.98	0.99	0.98	0.99	1.00	1.04	1.11	1.21	1.36
IPC	1.411	1.395	1.353	1.355	1.342	1.316		1.377	1.387	1.396	1.383	1.390	1.417	1.473	1.566	1.704	1.919
											Key:	<0.65	<0.75	<0.85	<0.95	<1.00	>1.00

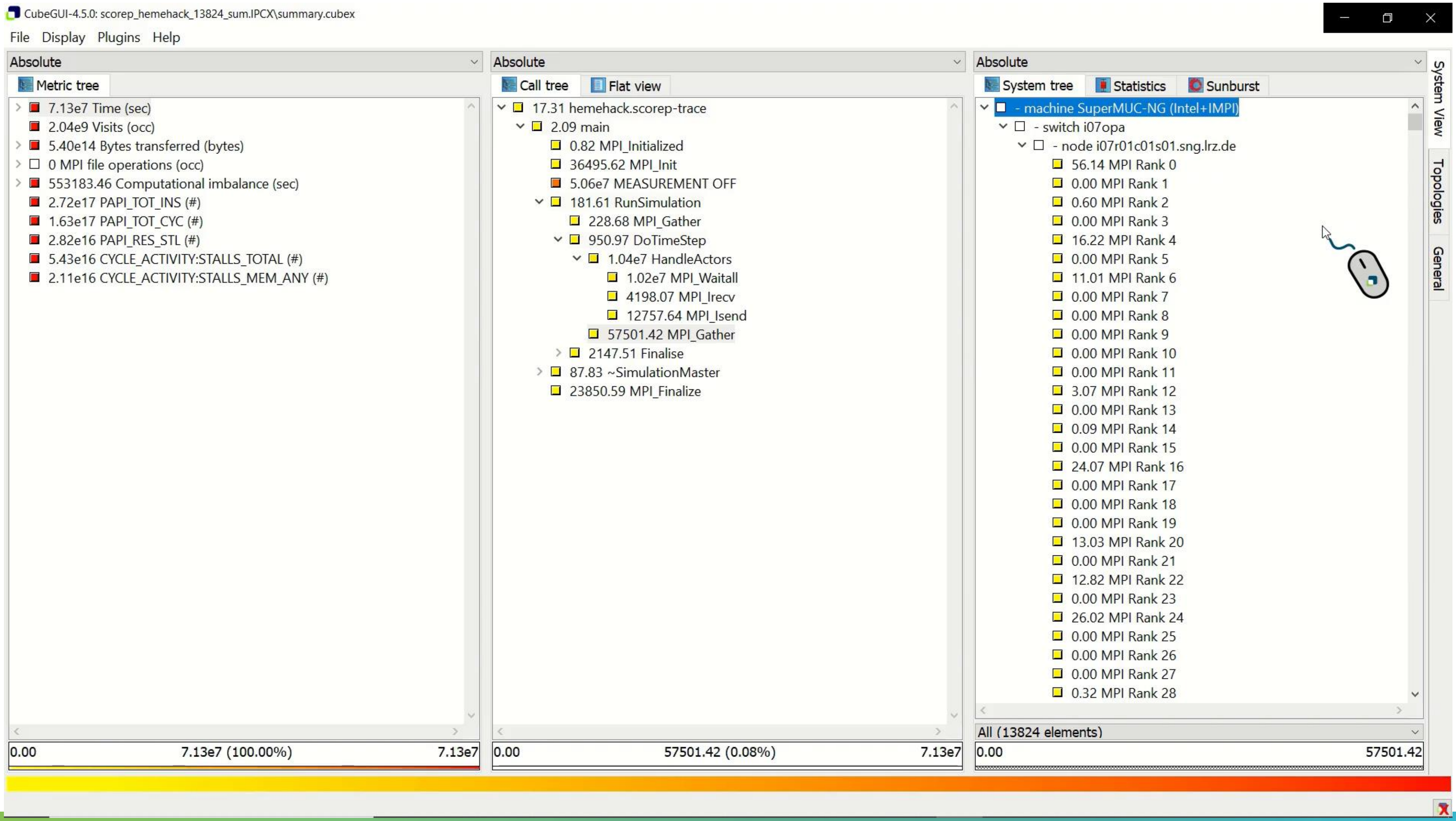
Global scaling efficiency fairly good around 80%, before degrading at larger scales

- Parallel efficiency deteriorating following Load balance efficiency
  - Communication efficiency excellent throughout
- Computation scaling (relative to 1152 processes) very good except at largest scale
  - Degradation of Instructions scaling partially compensated by improving IPC scaling

[POP CoE scaling efficiency model: www.pop-coe.eu]

Initial tree presentation: Time of MPI_Gather per MPI process

DOI 10.5281/zenodo.4080701



CubeGUI-4.5.0: scorep_hemehack_13824_sum.IPCX\summary.cubex

File Display Plugins Help

Absolute — Metric tree

```
> 7.13e7 Time (sec)
  2.04e9 Visits (occ)
> 5.40e14 Bytes transferred (bytes)
> 0 MPI file operations (occ)
> 553183.46 Computational imbalance (sec)
  2.72e17 PAPI_TOT_INS (#)
  1.63e17 PAPI_TOT_CYC (#)
  2.82e16 PAPI_RES_STL (#)
  5.43e16 CYCLE_ACTIVITY:STALLS_TOTAL (#)
  2.11e16 CYCLE_ACTIVITY:STALLS_MEM_ANY (#)
```

0.00 7.13e7 (100.00%) 7.13e7

Absolute — Call tree | Flat view

```
v 17.31 hemehack.scorep-trace
  v 2.09 main
      0.82 MPI_Initialized
      36495.62 MPI_Init
      5.06e7 MEASUREMENT OFF
    v 181.61 RunSimulation
        228.68 MPI_Gather
      v 950.97 DoTimeStep
        v 1.04e7 HandleActors
            1.02e7 MPI_Waitall
            4198.07 MPI_Irecv
            12757.64 MPI_Isend
          57501.42 MPI_Gather
      > 2147.51 Finalise
    > 87.83 ~SimulationMaster
      23850.59 MPI_Finalize
```

0.00 57501.42 (0.08%) 7.13e7

Absolute — System tree | Statistics | Sunburst

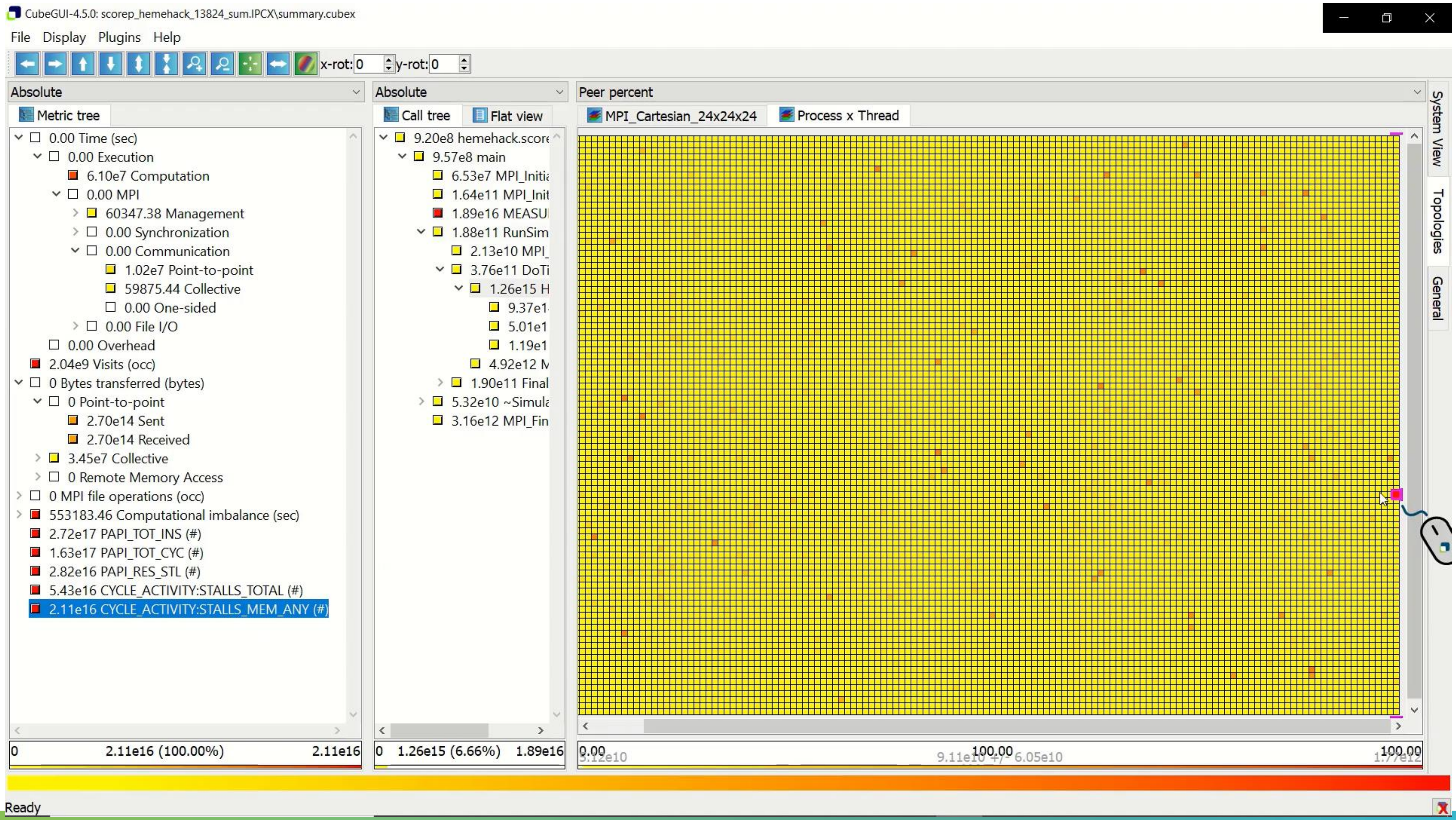
```
v - machine SuperMUC-NG (Intel+IMPI)
  v - switch i07opa
    v - node i07r01c01s01.sng.lrz.de
        56.14 MPI Rank 0
        0.00 MPI Rank 1
        0.60 MPI Rank 2
        0.00 MPI Rank 3
        16.22 MPI Rank 4
        0.00 MPI Rank 5
        11.01 MPI Rank 6
        0.00 MPI Rank 7
        0.00 MPI Rank 8
        0.00 MPI Rank 9
        0.00 MPI Rank 10
        0.00 MPI Rank 11
        3.07 MPI Rank 12
        0.00 MPI Rank 13
        0.09 MPI Rank 14
        0.00 MPI Rank 15
        24.07 MPI Rank 16
        0.00 MPI Rank 17
        0.00 MPI Rank 18
        0.00 MPI Rank 19
        13.03 MPI Rank 20
        0.00 MPI Rank 21
        12.82 MPI Rank 22
        0.00 MPI Rank 23
        26.02 MPI Rank 24
        0.00 MPI Rank 25
        0.00 MPI Rank 26
        0.00 MPI Rank 27
        0.32 MPI Rank 28
```

All (13824 elements)

0.00 57501.42

System View | Topologies | General

Topological presentation: STALLS_MEM_ANY for HandleActors



CubeGUI-4.5.0: scorep_hemehack_13824_sum.IPCX\summary.cubex

File Display Plugins Help

x-rot: 0 y-rot: 0

Absolute — Metric tree

```
0.00 Time (sec)
  0.00 Execution
    6.10e7 Computation
    0.00 MPI
      60347.38 Management
      0.00 Synchronization
      0.00 Communication
        1.02e7 Point-to-point
        59875.44 Collective
        0.00 One-sided
      0.00 File I/O
  0.00 Overhead
2.04e9 Visits (occ)
0 Bytes transferred (bytes)
  0 Point-to-point
    2.70e14 Sent
    2.70e14 Received
  3.45e7 Collective
  0 Remote Memory Access
0 MPI file operations (occ)
553183.46 Computational imbalance (sec)
2.72e17 PAPI_TOT_INS (#)
1.63e17 PAPI_TOT_CYC (#)
2.82e16 PAPI_RES_STL (#)
5.43e16 CYCLE_ACTIVITY:STALLS_TOTAL (#)
2.11e16 CYCLE_ACTIVITY:STALLS_MEM_ANY (#)
```

0 | 2.11e16 (100.00%) | 2.11e16

Absolute — Call tree / Flat view

```
9.20e8 hemehack.score
  9.57e8 main
    6.53e7 MPI_Initi
    1.64e11 MPI_Init
    1.89e16 MEASU
    1.88e11 RunSim
      2.13e10 MPI_
      3.76e11 DoTi
        1.26e15 H
          9.37e1
          5.01e1
          1.19e1
        4.92e12 M
      1.90e11 Final
    5.32e10 ~Simula
    3.16e12 MPI_Fin
```

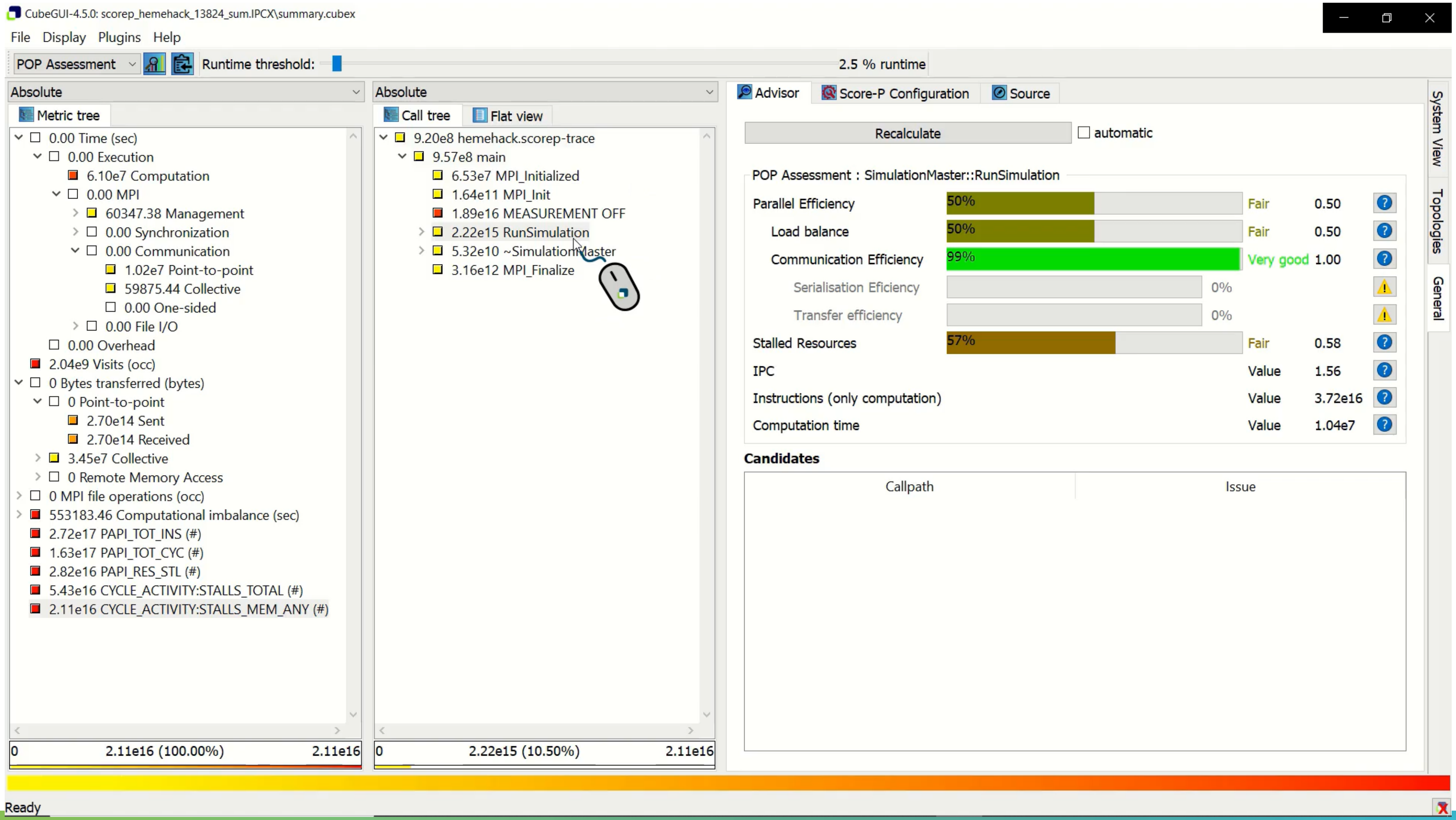
0 | 1.26e15 (6.66%) | 1.89e16

Peer percent — MPI_Cartesian_24x24x24 | Process x Thread

0.00 3.12e10 | 100.00 9.11e10 +/- 6.05e10 | 100.00 1.77e12

Ready

Advisor: POP efficiency assessment for RunSimulation



CubeGUI-4.5.0: scorep_hemehack_13824_sum.IPCX\summary.cubex

File Display Plugins Help

POP Assessment | Runtime threshold: 2.5 % runtime

Absolute — Metric tree

```
0.00 Time (sec)
  0.00 Execution
    6.10e7 Computation
    0.00 MPI
      60347.38 Management
      0.00 Synchronization
      0.00 Communication
        1.02e7 Point-to-point
        59875.44 Collective
        0.00 One-sided
      0.00 File I/O
  0.00 Overhead
2.04e9 Visits (occ)
0 Bytes transferred (bytes)
  0 Point-to-point
    2.70e14 Sent
    2.70e14 Received
  3.45e7 Collective
  0 Remote Memory Access
0 MPI file operations (occ)
553183.46 Computational imbalance (sec)
2.72e17 PAPI_TOT_INS (#)
1.63e17 PAPI_TOT_CYC (#)
2.82e16 PAPI_RES_STL (#)
5.43e16 CYCLE_ACTIVITY:STALLS_TOTAL (#)
2.11e16 CYCLE_ACTIVITY:STALLS_MEM_ANY (#)
```

0 2.11e16 (100.00%) 2.11e16

Absolute — Call tree | Flat view

```
9.20e8 hemehack.scorep-trace
  9.57e8 main
    6.53e7 MPI_Initialized
    1.64e11 MPI_Init
    1.89e16 MEASUREMENT OFF
    2.22e15 RunSimulation
    5.32e10 ~SimulationMaster
    3.16e12 MPI_Finalize
```

0 2.22e15 (10.50%) 2.11e16

Advisor | Score-P Configuration | Source

Recalculate ☐ automatic

POP Assessment : SimulationMaster::RunSimulation

Parallel Efficiency	50%	Fair	0.50
Load balance	50%	Fair	0.50
Communication Efficiency	99%	Very good	1.00
Serialisation Eficiency	0%		
Transfer efficiency	0%		
Stalled Resources	57%	Fair	0.58
IPC		Value	1.56
Instructions (only computation)		Value	3.72e16
Computation time		Value	1.04e7

Candidates

Callpath	Issue

System View | Topologies | General

Ready

HemeLB (JUWELS-Volta)

DOI 10.5281/zenodo.4117942

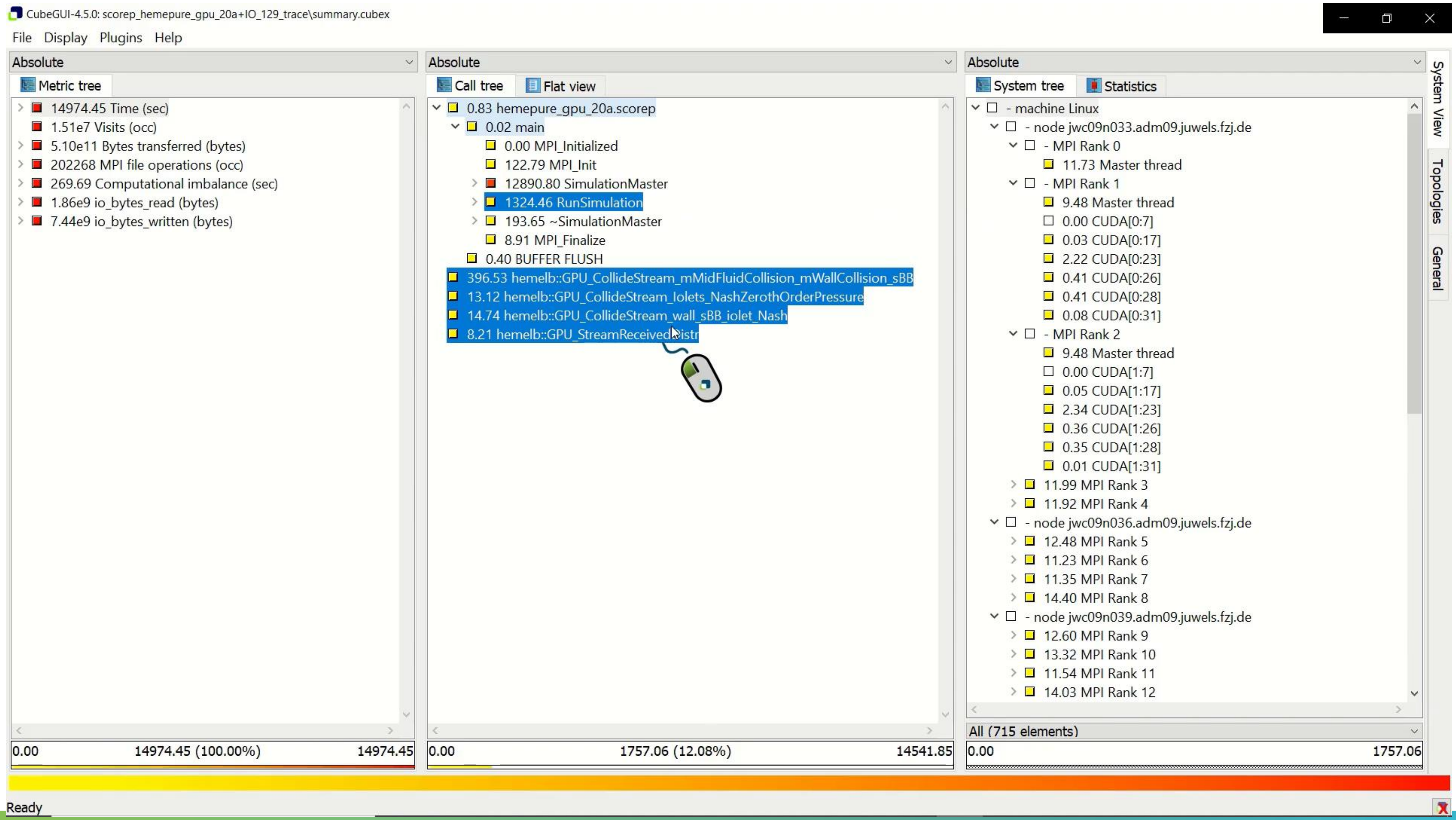
- 3D macroscopic blood flow in human arterial system developed by UC London (UK)
  - lattice-Boltzmann method tracking fluid particles on a lattice grid with complex boundary conditions
  - exascale flagship application of EU H2020 HPC Centre of Excellence for Computational Biomedicine
- HemeLB open-source code and test case: www.hemelb.org
  - C++ parallelized with MPI + CUDA (in development)
    - GCC/8.3.0 compiler, CUDA/10.1.105 and ParaStationMPI/5.4 library
    - configured with 2 'reader' processes and intermediate MPI file writing
    - rank 0 'monitor' process doesn't participate in simulation

  - Focus of analysis 2,000 time-step (each 100µs) simulation of CBM2019_Arteries_patched geometry
    - 1.78 GiB: 66,401,494 lattice sites, 1+38 iolets
- Executed on *JUWELS-Volta* (@JSC):
  - 2x 20-core Intel Xeon Platinum 8168 ('Skylake') CPUs + 4 Nvidia V100 'Volta' GPUs
  - 4* MPI processes/node (one per GPU), 32 (of 56) compute nodes: 129 MPI processes

⇒ ***Identification & quantification of impact of load balance and its variation***

Tree: Time for asynch. CUDA kernels on separate CUDA streams

DOI 10.5281/zenodo.4081080



CubeGUI-4.5.0: scorep_hemepure_gpu_20a+IO_129_trace\summary.cubex

File Display Plugins Help

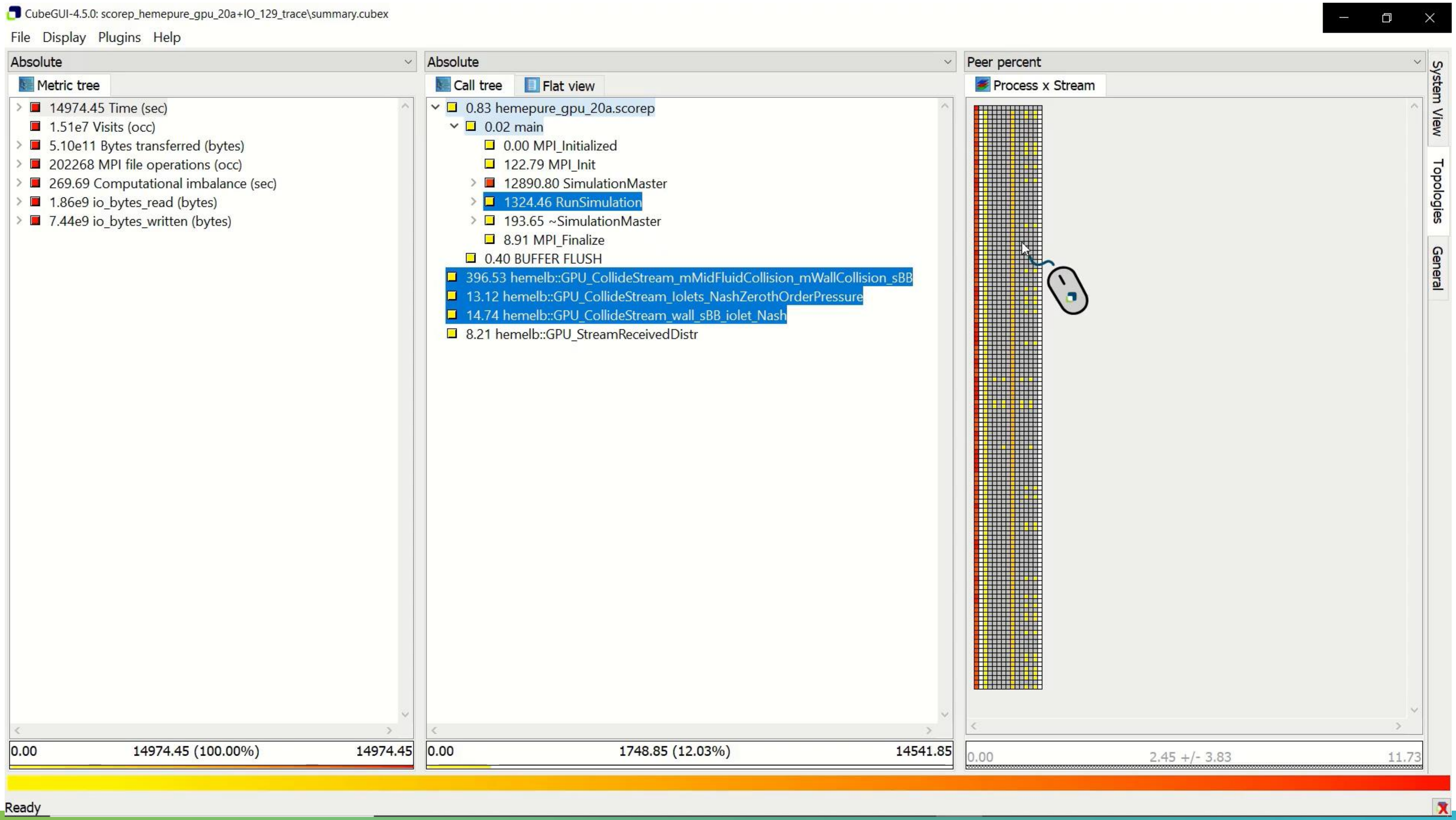
Absolute — Metric tree
```
> 14974.45 Time (sec)
  1.51e7 Visits (occ)
> 5.10e11 Bytes transferred (bytes)
> 202268 MPI file operations (occ)
> 269.69 Computational imbalance (sec)
> 1.86e9 io_bytes_read (bytes)
> 7.44e9 io_bytes_written (bytes)
```
0.00 | 14974.45 (100.00%) | 14974.45

Absolute — Call tree | Flat view
```
v 0.83 hemepure_gpu_20a.scorep
  v 0.02 main
      0.00 MPI_Initialized
      122.79 MPI_Init
    > 12890.80 SimulationMaster
    > 1324.46 RunSimulation
    > 193.65 ~SimulationMaster
      8.91 MPI_Finalize
    0.40 BUFFER FLUSH
  396.53 hemelb::GPU_CollideStream_mMidFluidCollision_mWallCollision_sBB
  13.12 hemelb::GPU_CollideStream_Iolets_NashZerothOrderPressure
  14.74 hemelb::GPU_CollideStream_wall_sBB_iolet_Nash
  8.21 hemelb::GPU_StreamReceivedDistr
```
0.00 | 1757.06 (12.08%) | 14541.85

Absolute — System tree | Statistics
```
v - machine Linux
  v - node jwc09n033.adm09.juwels.fzj.de
    v - MPI Rank 0
        11.73 Master thread
    v - MPI Rank 1
        9.48 Master thread
        0.00 CUDA[0:7]
        0.03 CUDA[0:17]
        2.22 CUDA[0:23]
        0.41 CUDA[0:26]
        0.41 CUDA[0:28]
        0.08 CUDA[0:31]
    v - MPI Rank 2
        9.48 Master thread
        0.00 CUDA[1:7]
        0.05 CUDA[1:17]
        2.34 CUDA[1:23]
        0.36 CUDA[1:26]
        0.35 CUDA[1:28]
        0.01 CUDA[1:31]
    > 11.99 MPI Rank 3
    > 11.92 MPI Rank 4
  v - node jwc09n036.adm09.juwels.fzj.de
    > 12.48 MPI Rank 5
    > 11.23 MPI Rank 6
    > 11.35 MPI Rank 7
    > 14.40 MPI Rank 8
  v - node jwc09n039.adm09.juwels.fzj.de
    > 12.60 MPI Rank 9
    > 13.32 MPI Rank 10
    > 11.54 MPI Rank 11
    > 14.03 MPI Rank 12
```
All (715 elements)
0.00 | 1757.06

Ready

Topo: Time for asynch. CUDA kernels on separate CUDA streams



CubeGUI-4.5.0: scorep_hemepure_gpu_20a+IO_129_trace\summary.cubex

File Display Plugins Help

Absolute — Metric tree
- 14974.45 Time (sec)
- 1.51e7 Visits (occ)
- 5.10e11 Bytes transferred (bytes)
- 202268 MPI file operations (occ)
- 269.69 Computational imbalance (sec)
- 1.86e9 io_bytes_read (bytes)
- 7.44e9 io_bytes_written (bytes)

0.00 | 14974.45 (100.00%) | 14974.45

Absolute — Call tree / Flat view
- 0.83 hemepure_gpu_20a.scorep
  - 0.02 main
    - 0.00 MPI_Initialized
    - 122.79 MPI_Init
    - 12890.80 SimulationMaster
    - 1324.46 RunSimulation
    - 193.65 ~SimulationMaster
    - 8.91 MPI_Finalize
  - 0.40 BUFFER FLUSH
- 396.53 hemelb::GPU_CollideStream_mMidFluidCollision_mWallCollision_sBB
- 13.12 hemelb::GPU_CollideStream_Iolets_NashZerothOrderPressure
- 14.74 hemelb::GPU_CollideStream_wall_sBB_iolet_Nash
- 8.21 hemelb::GPU_StreamReceivedDistr

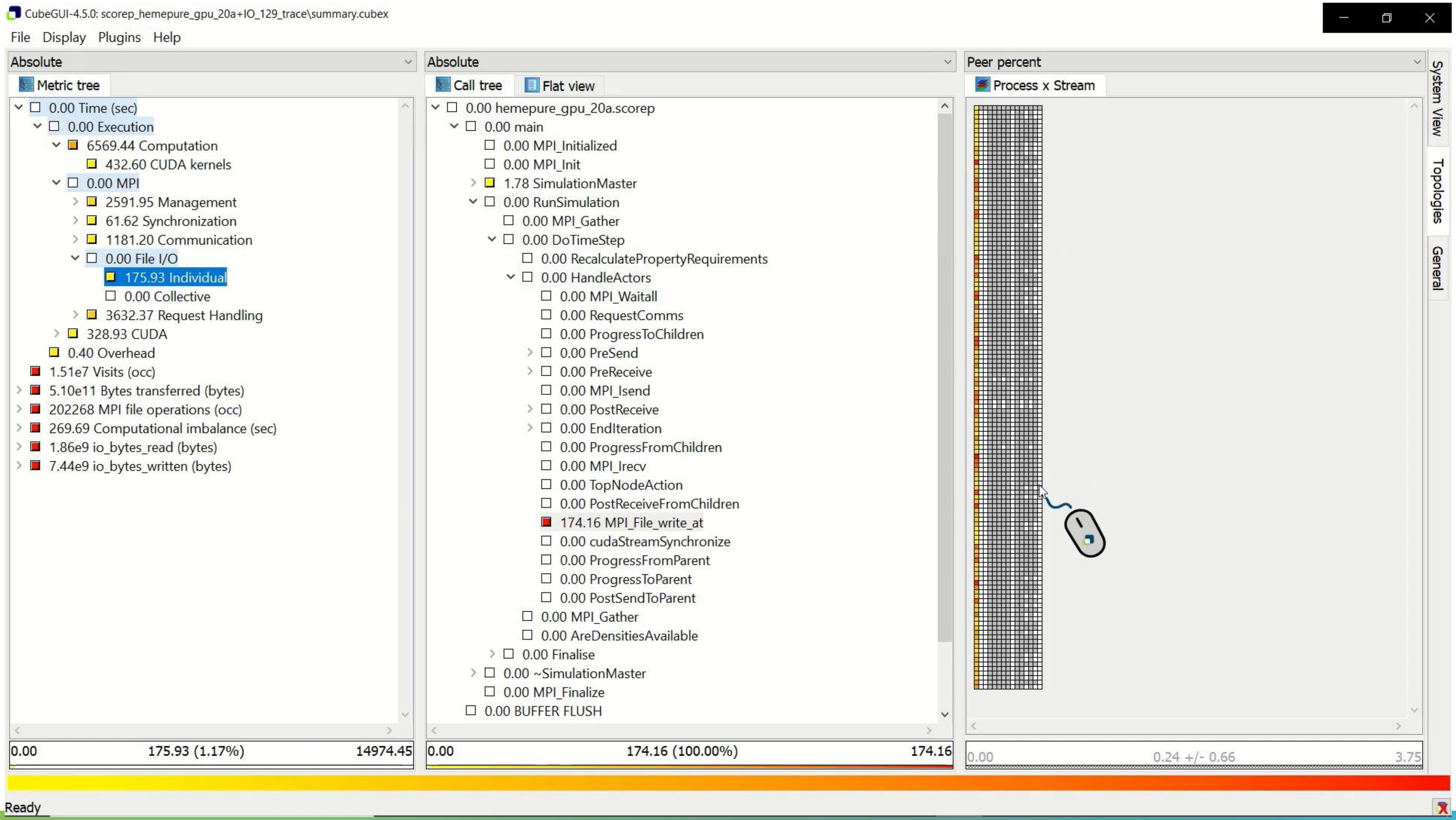
0.00 | 1748.85 (12.03%) | 14541.85

Peer percent — Process x Stream

0.00 | 2.45 +/- 3.83 | 11.73

Ready

Topo: Time for MPI file writing on CPU varies per MPI process


CubeGUI-4.5.0: scorep_hemepure_gpu_20a+IO_129_trace\summary.cubex

File Display Plugins Help

Absolute — Metric tree

```
0.00 Time (sec)
  0.00 Execution
    6569.44 Computation
      432.60 CUDA kernels
    0.00 MPI
      2591.95 Management
      61.62 Synchronization
      1181.20 Communication
      0.00 File I/O
        175.93 Individual
        0.00 Collective
      3632.37 Request Handling
    328.93 CUDA
  0.40 Overhead
1.51e7 Visits (occ)
5.10e11 Bytes transferred (bytes)
202268 MPI file operations (occ)
269.69 Computational imbalance (sec)
1.86e9 io_bytes_read (bytes)
7.44e9 io_bytes_written (bytes)
```

0.00 175.93 (1.17%) 14974.45

Absolute — Call tree / Flat view

```
0.00 hemepure_gpu_20a.scorep
  0.00 main
    0.00 MPI_Initialized
    0.00 MPI_Init
    1.78 SimulationMaster
    0.00 RunSimulation
      0.00 MPI_Gather
      0.00 DoTimeStep
        0.00 RecalculatePropertyRequirements
        0.00 HandleActors
          0.00 MPI_Waitall
          0.00 RequestComms
          0.00 ProgressToChildren
          0.00 PreSend
          0.00 PreReceive
          0.00 MPI_Isend
          0.00 PostReceive
          0.00 EndIteration
          0.00 ProgressFromChildren
          0.00 MPI_Irecv
          0.00 TopNodeAction
          0.00 PostReceiveFromChildren
          174.16 MPI_File_write_at
          0.00 cudaStreamSynchronize
          0.00 ProgressFromParent
          0.00 ProgressToParent
          0.00 PostSendToParent
        0.00 MPI_Gather
        0.00 AreDensitiesAvailable
      0.00 Finalise
    0.00 ~SimulationMaster
    0.00 MPI_Finalize
  0.00 BUFFER FLUSH
```

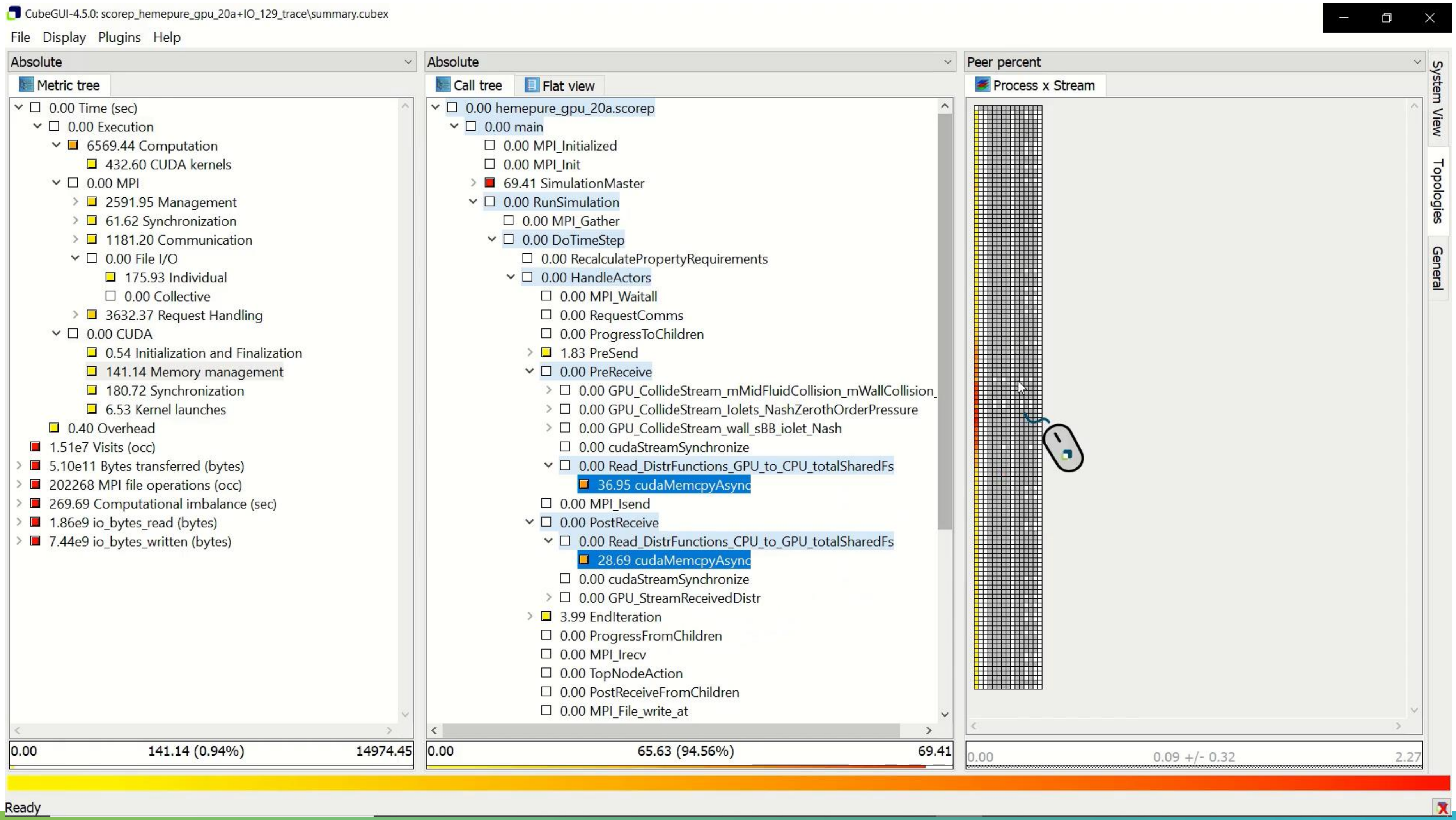
0.00 174.16 (100.00%) 174.16

Peer percent — Process x Stream

0.00 0.24 +/- 0.66 3.75

Ready

Topo: Time for CUDA asynchronous memory copies is imbalanced



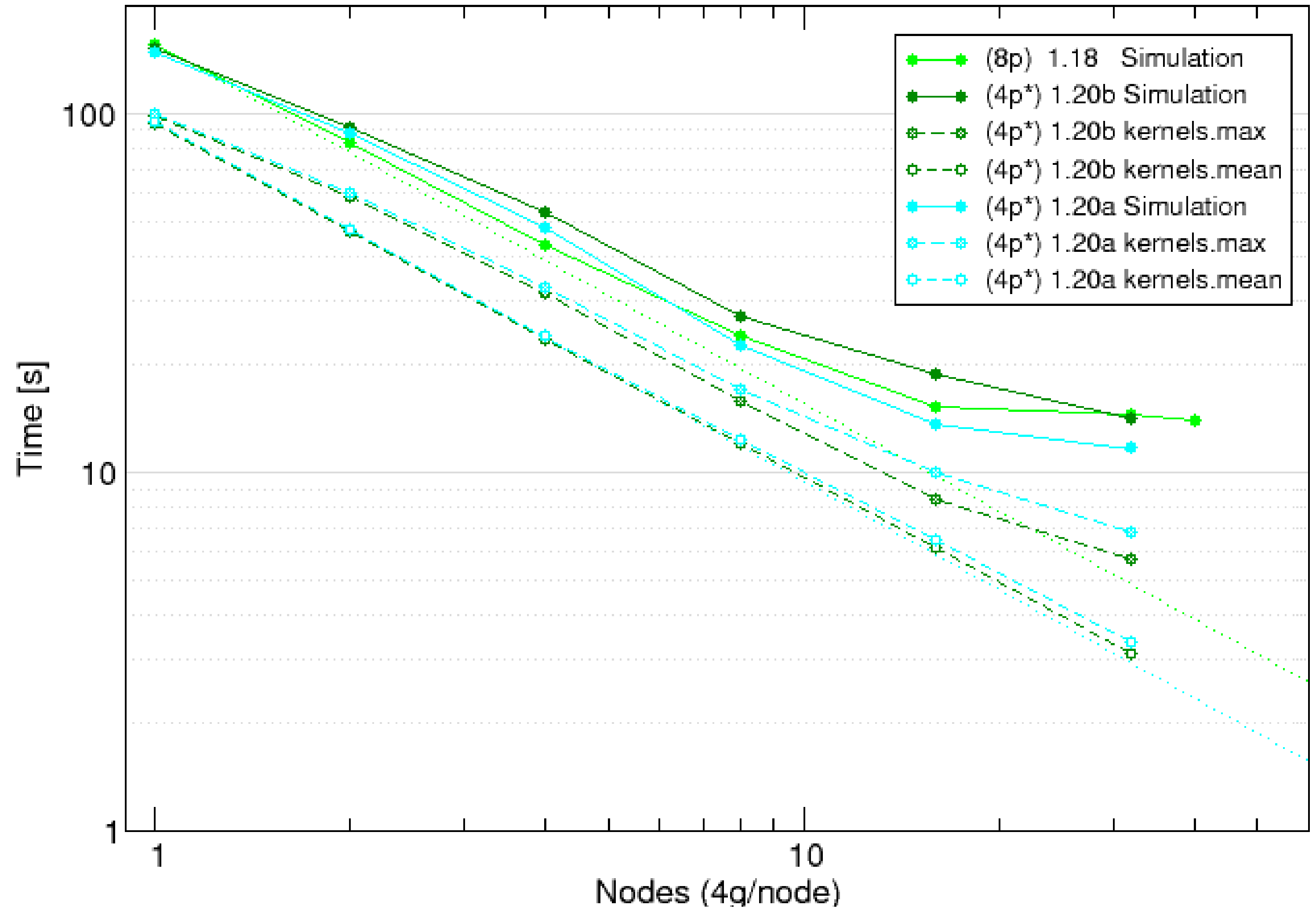
CubeGUI-4.5.0: scorep_hemepure_gpu_20a+IO_129_trace\summary.cubex

```
0.00 Time (sec)
  0.00 Execution
    6569.44 Computation
      432.60 CUDA kernels
    0.00 MPI
      2591.95 Management
      61.62 Synchronization
      1181.20 Communication
      0.00 File I/O
        175.93 Individual
        0.00 Collective
      3632.37 Request Handling
    0.00 CUDA
      0.54 Initialization and Finalization
      141.14 Memory management
      180.72 Synchronization
      6.53 Kernel launches
  0.40 Overhead
1.51e7 Visits (occ)
5.10e11 Bytes transferred (bytes)
202268 MPI file operations (occ)
269.69 Computational imbalance (sec)
1.86e9 io_bytes_read (bytes)
7.44e9 io_bytes_written (bytes)

0.00    141.14 (0.94%)    14974.45
```

```
0.00 hemepure_gpu_20a.scorep
  0.00 main
    0.00 MPI_Initialized
    0.00 MPI_Init
    69.41 SimulationMaster
    0.00 RunSimulation
      0.00 MPI_Gather
      0.00 DoTimeStep
        0.00 RecalculatePropertyRequirements
        0.00 HandleActors
          0.00 MPI_Waitall
          0.00 RequestComms
          0.00 ProgressToChildren
          1.83 PreSend
          0.00 PreReceive
            0.00 GPU_CollideStream_mMidFluidCollision_mWallCollision_
            0.00 GPU_CollideStream_Iolets_NashZerothOrderPressure
            0.00 GPU_CollideStream_wall_sBB_iolet_Nash
            0.00 cudaStreamSynchronize
            0.00 Read_DistrFunctions_GPU_to_CPU_totalSharedFs
              36.95 cudaMemcpyAsync
          0.00 MPI_Isend
          0.00 PostReceive
            0.00 Read_DistrFunctions_CPU_to_GPU_totalSharedFs
              28.69 cudaMemcpyAsync
            0.00 cudaStreamSynchronize
            0.00 GPU_StreamReceivedDistr
          3.99 EndIteration
          0.00 ProgressFromChildren
          0.00 MPI_Irecv
          0.00 TopNodeAction
          0.00 PostReceiveFromChildren
          0.00 MPI_File_write_at

0.00    65.63 (94.56%)    69.41
```

```
0.00    0.09 +/- 0.32    2.27
```

HemeLB@JUWELS-Volta strong scaling of FOA *RunSimulation*



- Reference execution with 8ppn
  - multiple processes offloading GPU kernels generally unproductive
- Comparison of versions (4ppn)
  - v1.20a generally better
- Synchronous MPI file writing is the primary bottleneck
- CUDA kernels on GPUs
  - less than half of Simulation time (therefore GPUs mostly idle)
  - total kernel time scales very well (0.93 scaling efficiency)
  - load balance deteriorates (0.95 for single node, 0.50 for 32 nodes)

HemeLB@JUWELS/Volta strong scaling efficiency of *RunSimulation*

	1n 5p	2n 9p	4n 17p	8n 33p	16n 65p	32n 129p
Simulation time [s]	147.87	88.38	48.13	22.66	13.68	11.67
Global scaling efficiency	0.64	0.53	0.49	0.52	0.43	0.25
– Parallel efficiency	0.64	0.53	0.50	0.54	0.47	0.29
– – Load balance efficiency (GPU)	0.95	0.78	0.73	0.73	0.65	0.50
– – Communication efficiency (GPU)	0.67	0.68	0.68	0.75	0.73	0.58
– Computation scaling (GPU)	1.00	1.00	0.99	0.96	0.92	0.87

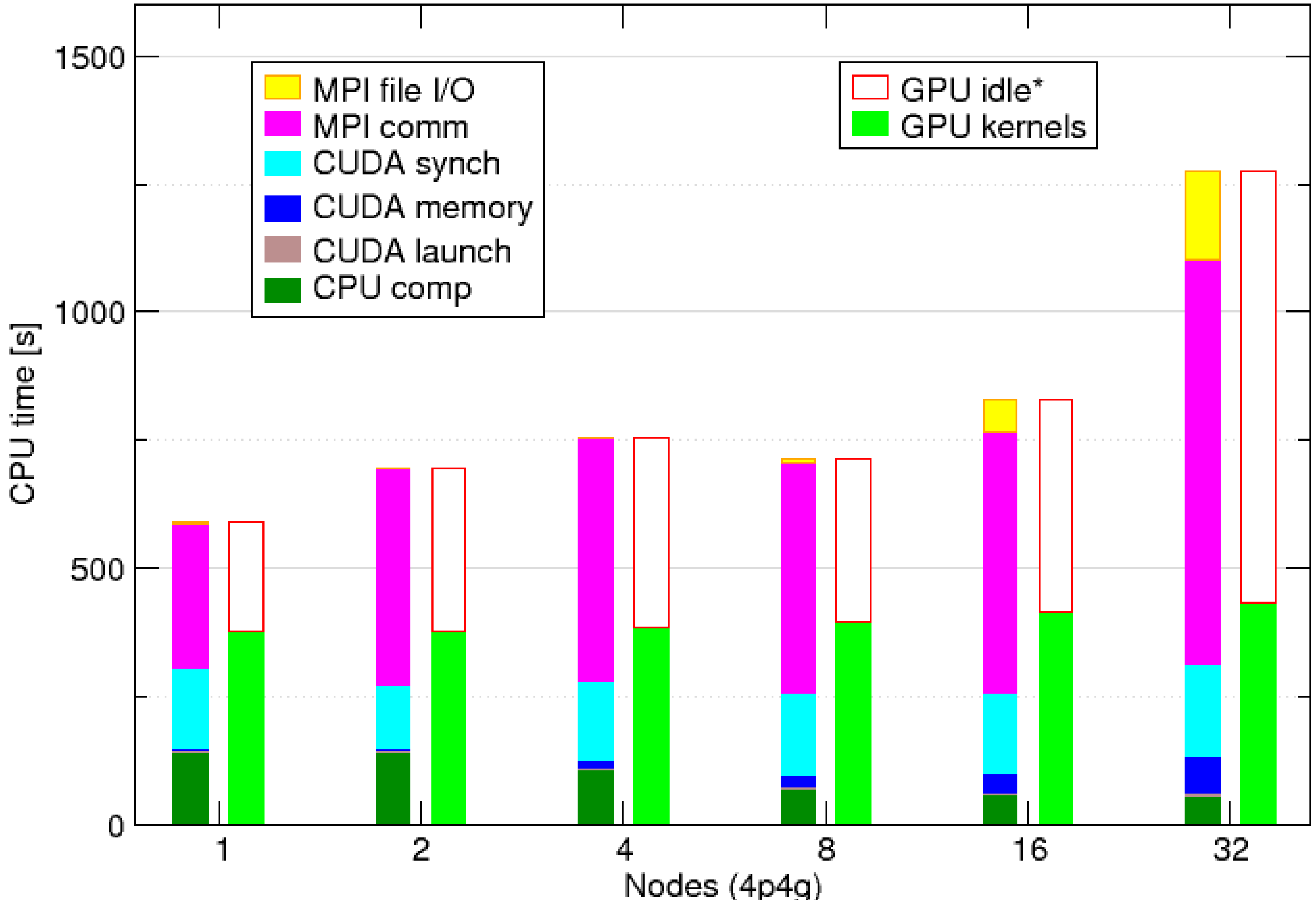
Key:
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Only considering GPUs (ignoring all CPU cores, 90% of which are completely unused)

- Single (quad-GPU) node already suffers significant communication inefficiency
  - includes MPI file writing, but doesn't degrade much as additional nodes are included
- Load balance of GPUs deteriorates progressively
- GPU computation scaling remains reasonably good

[POP CoE scaling efficiency model: www.pop-coe.eu]

HemeLB@JUWELS-Volta strong scaling of FOA *RunSimulation*



- CPU+GPU time breakdown
- CUDA kernels on GPUs
  - less than half of Simulation time (therefore GPUs mostly idle)
  - total kernel time scales very well (0.87 scaling efficiency)
- MPI processes on CPUs
  - computation time decreases
  - CUDA synchronization time fairly constant, but time for memory management increases somewhat
  - MPI communication time dominates, with much more time for file writing with 16+ nodes