TECHNISCHE UNIVERSITÄT DRESDEN

Center for Information Services and High Performance Computing (ZIH)

Analyzing Parallel I/O BOF:

Vampir(Trace)

Zellescher Weg 12

Willers-Bau A 208

Tel. +49 351 - 463 – 34217

Michael Kluge (michael.kluge@tu-dresden.de)

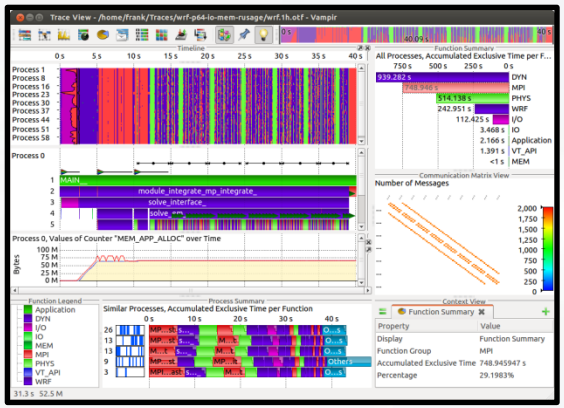
ZIH
Center for Information Services & High Performance Computing

Vampir Mission

Visualization of dynamics
of complex parallel processes

Requires two components

- Monitor/Collector (VampirTrace or Score-P)
- Charts/Browser (Vampir)



Typical questions that Vampir helps to answer:

- What happens in my application execution during a given time in a given process or thread?
- How do the communication patterns of my application execute on a real system?
- Are there any imbalances in computation, I/O or memory usage and how do they affect the parallel execution of my application?

Performance Tools

- Community efforts to make tools more versatile

Measurement: **Score-P**

Partners:

Technische Universität München

gns

JÜLICH FORSCHUNGSZENTRUM

TECHNISCHE UNIVERSITÄT DRESDEN

RWTH AACHEN UNIVERSITY

German Research School for Simulation Sciences

UNIVERSITY OF OREGON

Analysis:

- TAU

Tuning and Analysis Utilities

- **Vampir**

VAMPIR

- Scalasca

scalasca

Tracing – Monitors

- Currently for I/O VampirTrace is still the better tracing tool

Measurement: **Score-P**

Partners:

Technische Universität München

gns

JÜLICH FORSCHUNGSZENTRUM

TECHNISCHE UNIVERSITÄT DRESDEN

RWTH AACHEN UNIVERSITY

German Research School for Simulation Sciences

UNIVERSITY OF OREGON

Measurement:
VampirTrace
(Legacy)
Just for Vampir

Analysis:

- TAU

Tuning and Analysis Utilities

- **Vampir**

VAMPIR

- Scalasca

scalasca

Roadmap

November 2014 | June 2015 | Longer term

Vampir:

- Collapsing timelines

Score-P:

- Sampling+tracing
- Xeon Phi (Native+Offload)
- Full SIONlib support (multi-flush, MPI+OpenMP)
- Memory tracing
- CPU ID
- 3rd party library wrapping
- OpenCL support

Score-P:

- Cobi (Dyninst) instrumentation
- **I/O tracing**
- More MPI-3 support

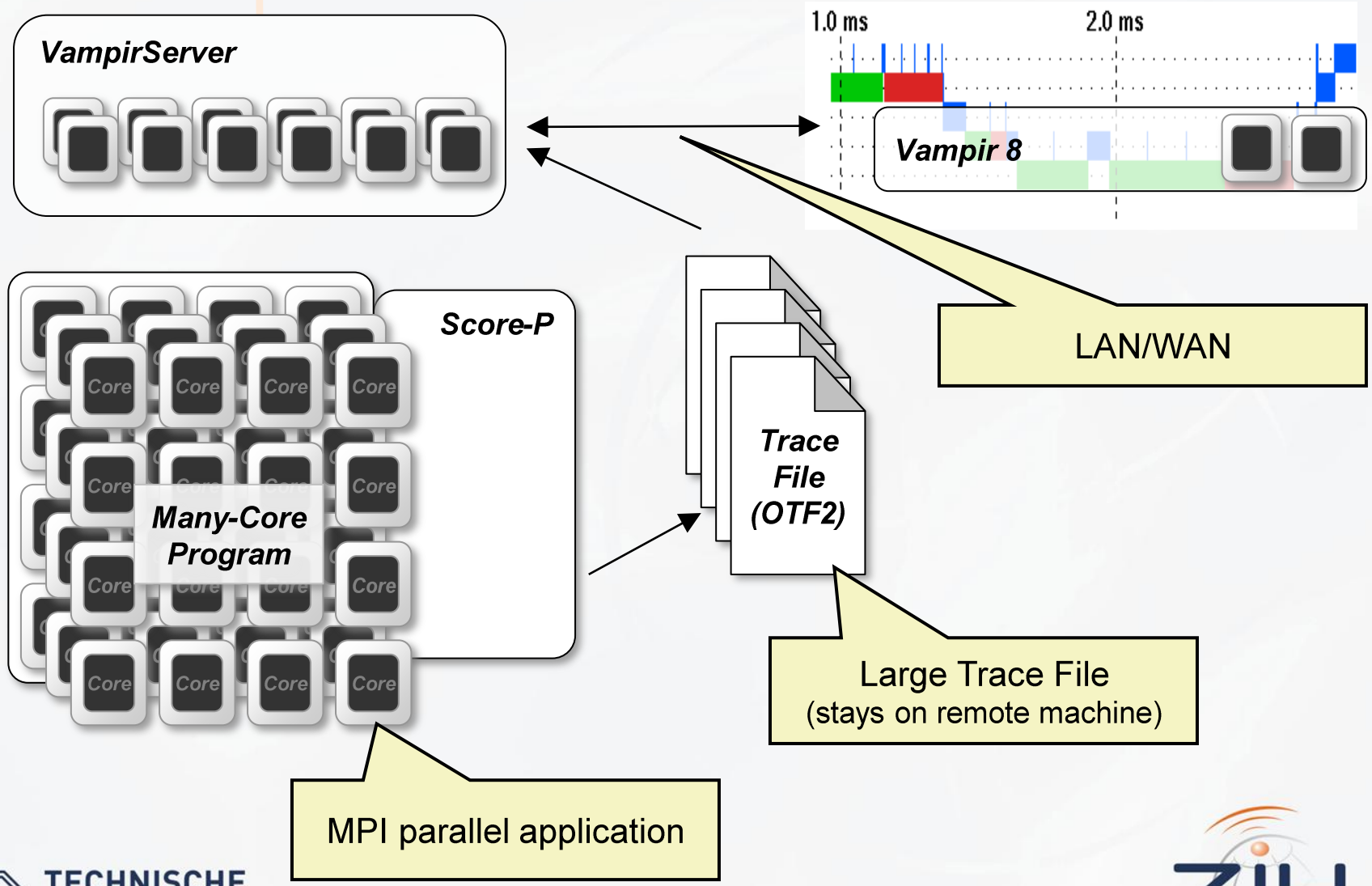
Vampir:

- Evaluation of traces with semantic compression

Score-P:

- Generation of traces with semantic compression

Vampir Workflow



Vampir I/O Analysis

- Load the trace file -> Information about all I/O accesses in memory
- Analyze via the GUI:
  - take: number of operations, operation size, operation time, I/O bandwidth (min/max/avg)
  - group by file name, access size or operation type
  - filter by process id or I/O operation
  - sort
  - graph

Vampir & I/O

Trace View – /Users/holger/Archive/traces/2009-10-02_lieber_wrf-io/wrf.otf – Vampir

5.4631 s – 5.4646 s
1.459 ms

Timeline

5.4635 s | 5.4640 s | 5.4645 s

Process 0
Process 1 — MPI_Bcast
Process 2 — MPI_Bcast
Process 3 — MPI_Bcast
Process 4 — MPI_Bcast
Process 5 — MPI_Bcast
Process 6 — MPI_Bcast
Process 7 — MPI_Bcast
Process 8 — MPI_Bcast
Process 9 — MPI_Bcast
Process 10 — MPI_Bcast

Process 0

1 WRF
2 MED_INITIALDATA_INPUT
3 MODULE_IO_DOMAIN::OPEN_R_DATASET
4 MODULE_IO::WRF_OPEN_FOR_READ
5 EXT_NCD_OPEN_FOR_READ
6 EXT_NCD_OPEN_FOR_READ_BEGIN
7 malloc
8

Function Summary

All Processes, Accumulated Exclusive Tim...

50 µs | 25 µs | 0 µs

69.45 µs open
50.55 µs read
7.8 µs lseek

Context View

Master Timeline

Property	Value
Display	Master Timeline
Type	I/O Event
Initiator	Process 0
File Name	wrfinput_d01
File Group	fileio
Operation	READ
Interval Begin	5.463816 s
Interval End	5.463866 s
Duration	50.55 µs
Size	16 KiB
Bandwidth	309.099901 MiB/s

5.46315 s – 5.46455 s (1.4 ms)

Vampir & I/O

Trace View – /Users/holger/Archive/traces/2009-10-02_lieber_wrf-io/wrf.otf – Vampir

0 s 39.042 s 39 s

Timeline

0 s 10 s 20 s 30 s

Process 0
Process 1
Process 2
Process 3
Process 4
Process 5
Process 6
Process 7
Process 8
Process 9
Process 10
Process 11

Function Summary

All Processes, Accumulated Exclusive Tim...

0.1 s 0.0 s

0.213 s	open
0.131 s	read
79.805 ms	write
35.145 ms	lseek
3.832 ms	open64
2.451 ms	fopen
<1 ms	dup2
<1 ms	close
<1 ms	fscanf
<1 ms	fclose
<100 µs	lseek64

I/O Summary

All Processes, Average I/O Bandwidth per File Name

1,280 MiB/s 640 MiB/s 0 MiB/s

1.879 GiB/s	RRTM_DATA
1.76 GiB/s	wrfbdy_d01
1.479 GiB/s	wrfinput_d01
730.14 MiB/s	namelist.input
633.874 MiB/s	LANDUSE.TBL
38.24 KiB/s	/proc/sgi_sn/...serial_number

Vampir I/O Analysis

Trace View – /home/mkluge/trace_roscon_mpi_hdf5_python/python_mpi_hfd5_1.otf * – Vampir

File Edit Chart Filter Window Help

41.735 s

Timeline

0 s 15 s 30 s

Process 0
Process 15
Process 29
Process 43
Process 57
Process 72
Process 87
Process 101
Process 115
Process 130
Process 145
Process 159
Process 173
Process 187
Process 202
Process 217
Process 231
Process 245
Process 261
Process 275
Process 291
Process 305
Process 321
Process 336
Process 350
Process 365
Process 379
Process 393

I/O Summary

All Processes, Aggregated I/O Transaction Size per File N...

80 GiB	0 GiB
119.209 GiB	minimal_example_output.hdf5
107.368 MiB	/scratch/mkluge...add_newdocs.pyc
86.226 MiB	/scratch/mkluge...mpy/ma/core.pyc
44.192 MiB	/scratch/mkluge/...unction_base.pyc
42.884 MiB	<UNKNOWN FILE DESCRIPTOR>
37.536 MiB	/etc/hosts
36.452 MiB	/scratch/mkluge/.../fromnumeric.pyc
33.333 MiB	/scratch/mkluge/...core/numeric.pyc
30.648 MiB	/scratch/mkluge/...defchararray.pyc
26.389 MiB	/scratch/mkluge/p...linalg/linalg.pyc
25.523 MiB	/scratch/mkluge/...al/chebyshev.pyc
24.432 MiB	/scratch/mkluge/p...on2.7/difflib.pyc
22.755 MiB	/scratch/mkluge/...ial/legendre.pyc
22.71 MiB	/scratch/mkluge/...al/hermite_e.pyc
22.674 MiB	/scratch/mkluge/...mial/hermite.pyc
22.524 MiB	/scratch/mkluge/p...testing/utils.pyc
22.466 MiB	/scratch/mkluge/...ial/laguerre.pyc
21.343 MiB	/scratch/mkluge/...py/lib/npyio.pyc
21.214 MiB	/scratch/mkluge/...py/ma/extras.pyc
19.537 MiB	/scratch/mkluge/...l/polynomial.pyc

Vampir I/O Analysis

Trace View – /home/mkluge/trace_roscon_mpi_hdf5_python/python_mpi_hfd5_1.otf * – Vampir

File Edit Chart Filter Window Help

41.735 s

Timeline

0 s 15 s 30 s

Process 0
Process 15
Process 29
Process 43
Process 57
Process 72
Process 87
Process 101
Process 115
Process 130
Process 145
Process 159
Process 173
Process 187
Process 202
Process 217
Process 231
Process 245
Process 261
Process 275
Process 291
Process 305
Process 321
Process 336
Process 350
Process 365
Process 379
Process 393

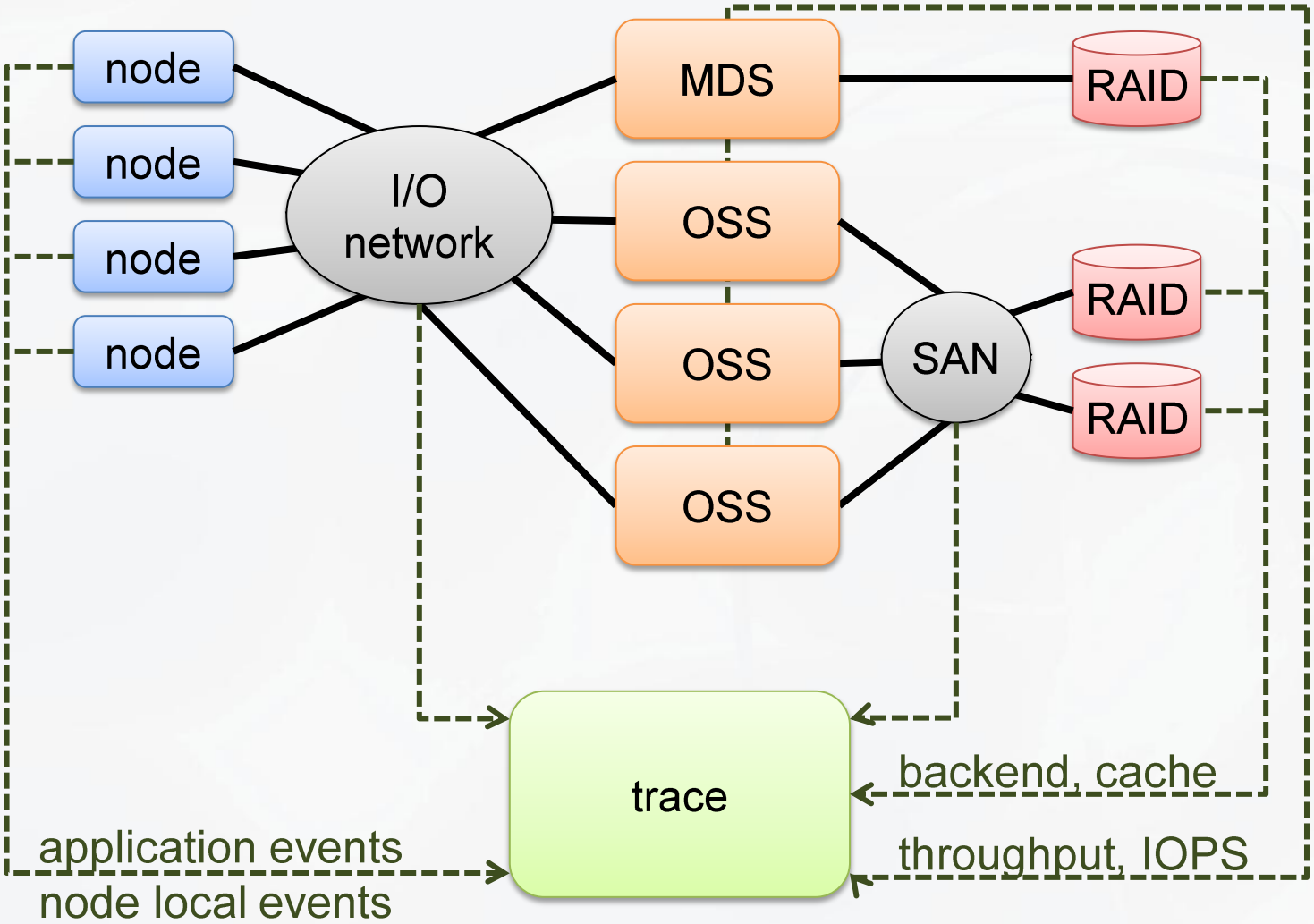
I/O Summary

All Processes, Number of I/O Operations per File Name

750 k 500 k 250 k 0 k

840,164	<UNKNOWN ...SCRIPTOR>
295,600	/scratch/mk...newdocs.pyc
152,321	/sw/taurus/li...hc-rules.conf
122,800	/proc/cpuinfo
28,400	<STDIN>
21,508	/dev/infiniband/uverbs0
16,829	/sw/taurus/l..._dpm_orte.la
16,400	/sw/taurus/li...a_io_romio.la
16,400	/sw/taurus/l...bsub_orte.la
16,400	/sw/taurus/li...nity_hwloc.la
16,400	/sw/taurus/li..._first_use.la
16,400	/sw/taurus/l...opo_slurm.la
16,400	/sw/taurus/l...po_openib.la
16,400	/sw/taurus/l...opo_hwloc.la
16,400	/sw/taurus/l...thostname.la
16,400	/sw/taurus/l...opo_cpuid.la
16,400	/sw/taurus/li...po_bullbcs.la
16,400	/sw/taurus/l..._osc_rdma.la
16,400	/sw/taurus/li..._osc_pt2pt.la
16,400	/sw/taurus/l...ca_ghc_sm.la

System Interaction (1)



Interaction:

- MPI
- I/O calls
- I/O system

System Interaction (2)

Vampir - [Trace View - /home/mkluge/threader_scaling_trace/dios_mpi_23.otf]

File Edit Chart Filter Window Help

Timeline

100 s 200 s 300 s 400 s 500 s 600 s

Process 28 Process 29 Process 30 Process 31 Process 32

MPI_Barrier

ExtCounter 0:1, Values of Counter "MDS001 remote procedure calls/s" over Time

10 k 5 k 0 k

metadata rate

ExtCounter 0:1, Values of Counter "pc_san: read bandwidth" over Time

44.25 M 43.50 M 42.75 M 42.00 M

metadata server load

ExtCounter 0:1, Values of Counter "mds001/load_1" over Time

4 k 3 k 2 k 1 k

metadata server load

Function Summary

All Processes, Accumulated Exclusiv...

10,000 s

31,116.835 s fopen64
29,824.819 s fclose
12,937.123 s MPI_Barrier
10,830.98 s fread

Function Legend

Application
LIBC-I/O
MPI
VT_API

I/O Summary

All Processes, Read, Sum of Size per File

0 GiB

73.626 GiB Sum
14.305 MiB /fastfs/mkluge/threader_database_single/164
14.301 MiB /fastfs/mkluge/threader_database_single/3768
14.301 MiB /fastfs/mkluge/threader_database_single/2696

Context View

Master Timeline

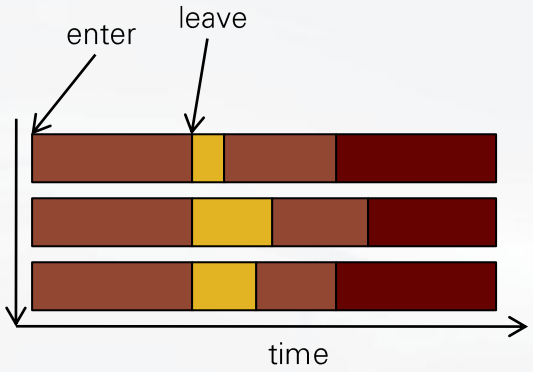
Property	Value
Display	Master Timeline
Type	I/O Event
Initiator	Process 5
File Name	/fastfs/mkluge/threader_database_single/3234
File Group	fileio
Operation	READ
Interval Begin	252.270591 s
Interval End	252.335219 s
Duration	64.6285 ms
Size	21.536133 KiB
Data Rate	333.229654 KiB/s

Interaction:

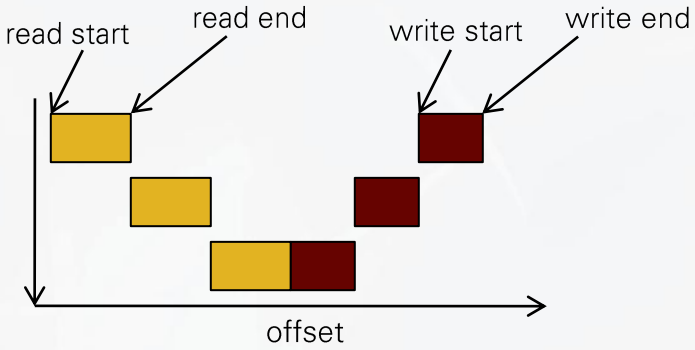
- MPI
- I/O calls
- I/O system

File Access Pattern Visualization

- Standard Timeline shows:
  - X axis: time
  - Y axis: processes



- What is needed for file access patterns:
  - X axis: byte range within a single file
  - Y axis: processes



- Possibilities:
  - mark multiple writes, read-after-write conditions, etc.

Access Pattern Visualization

Vampir - [Trace View - access.otf]

File View Help

View Chart Filter

Timeline

1400000000 s 1600000000 s 1800000000 s

Process 5
Process 14
Process 23
Process 32
Process 41
Process 50
Process 59
Process 68
Process 77
Process 86
Process 95
Process 104
Process 113
Process 122

Context View

Master Timeline

Property	Value
Display	Master Timeline
Type	Function
Function Name	write
Function Group	2
Interval Begin	1776950820 s
Interval End	1778597412 s
Duration	1646592 s
Source File	
Source Line	

Access Pattern Visualization

Vampir - [Trace View - access.otf]

File View Help

View Chart Filter

Timeline

2550000 s 2600000 s 2650000 s 2700000 s

Process 40
Process 42
Process 44
Process 46
Process 48
Process 50
Process 52
Process 54
Process 56
Process 58
Process 60
Process 62
Process 64
Process 66
Process 68
Process 70

Context View

Master Timeline

Property	Value
Display	Master Timeline
Type	Function
Function Name	write
Function Group	2
Interval Begin	2570116 s
Interval End	2574644 s
Duration	4528 s
Source File	
Source Line	