

Barcelona Supercomputing Center Centro Nacional de Supercomputación



Judit Giménez, <u>Lau Mercadal</u>, Germán Llort

Extrae

&

Paraver



Performance Analysis and Tools

2024-09-05

Extrae features

- Platforms
 - Intel, ARM, RISC-V, POWER, Cray, BlueGene, Android, Fujitsu Sparc ...
- Parallel programming models MPI, OpenMP, pthreads, OmpSs, CUDA, OpenACC, Java, Python ... • Performance Counters No need Using PAPI interface to Link to source code recompile nor relink! Callstack at MPI routines OpenMP outlined routines Selected user functions (Dyninst) Periodic sampling
- User events anywhere in your program (Extrae API)



Extrae overhead

	Karolina
Punctual event	175 ns
Event + PAPI counters	319 ns
Event + 1-level callstack	1.471 us
Event + 6-levels callstack	3.461 us



How does Extrae work?

• Symbol substitution through LD_PRELOAD

export LD_PRELOAD=\$EXTRAE_HOME/lib/libmpitrace.so

- Specific libraries for each runtime and combinations
 - MPI
 - OpenMP
 - OpenMP+MPI
 - CUDA
 - ...
- Dynamic instrumentation
 - Based on Dyninst (developed by U.Wisconsin / U.Maryland)
 - Instrumentation in memory
 - Binary rewriting

• Static link (i.e., PMPI, Extrae API)



Recommended

Using Extrae in 3 steps

- **1.** Adapt your job submission scripts
- 2. Configure what to trace
 - XML configuration file
 - Example configurations at \$EXTRAE_HOME/share/example
- **3. Run** it!

- For further reference check the Extrae User Guide:
 - <u>https://tools.bsc.es/doc/html/extrae</u>
 - Also distributed with Extrae at \$EXTRAE_HOME/share/doc



Step 1: Adapt the job script to load Extrae

hpc**\$ vi job.slur**m





Step 1: Adapt the job script to load Extrae

hpc\$ vi trace.sh





Step 1: Which tracing library?

• Choose depending on the application type

Library	Serial	MPI	OpenMP	pthread	CUDA
libseqtrace	\checkmark				
libmpitrace[f] ¹		\checkmark			
libomptrace			\checkmark		
libpttrace				\checkmark	
libcudatrace					\checkmark
libompitrace[f] ¹		\checkmark	\checkmark		
libptmpitrace[f] ¹		\checkmark		\checkmark	
libcudampitrace[f] ¹		\checkmark			\checkmark

¹ include suffix "f" in Fortran codes



Step 3: Run it!

• Submit your job

hpc\$ sbatch job.slurm

• Easy!



Step 2: Extrae XML configuration

hpc\$ vi extrae.xml





Step 2: Extrae XML configuration (II)

hpc\$ vi extrae.xml

```
<counters enabled="yes">
 <cpu enabled="yes" starting-set-distribution="1">
   <set enabled="ves" domain="all" changeat-time="500000us">
     PAPI TOT INS, PAPI TOT CYC, PAPI L1 DCM, PAPI L2 DCM,
     PAPI L3 TCM, PAPI BR INS, PAPI BR MSP, RESOURCE STALLS
   </set>
   <set enabled="yes" domain="all" changeat-time="500000us">
     PAPI TOT INS, PAPI TOT CYC, PAPI VEC SP, PAPI SR INS, PAPI LD INS
   </set>
 </cpu>
                                                              Select which
 <network enabled="no" />
 <resource-usage enabled="no" />
                                                              HW counters
 <memory-usage enabled="no" />
                                                              are measured
</counters>
                                                         (How's the machine doing?)
```



Step 2: Extrae XML configuration (III)

hpc\$ vi extrae.xml





All done! Check your resulting trace

• Once finished (check with "squeue") you will have the trace (3 files):

```
hpc$ ls -1
....
lulesh2.0_s65_27p.pcf
lulesh2.0_s65_27p.prv
lulesh2.0_s65_27p.row
```

• Now let's look into it !





Barcelona Supercomputing Center Centro Nacional de Supercomputación



Paraver Installation

Judit Giménez, Lau Mercadal, Germán Llort



2024-09-05

PATC: Performance Analysis and Tools

Install Paraver

Download from https://tools.bsc.es/downloads





Install Tutorials (Automatic)

Start Paraver

laptop\$ paraver/bin/wxparaver





Install Tutorials (Manual)

- Download tutorials:
 - Documentation \rightarrow Paraver Tutorials



laptop\$ mkdir paraver/tutorials

laptop\$ tar xf 3.introduction* -C paraver/tutorials/



Install Tutorials (Manual)

Start Paraver

laptop\$ paraver/bin/wxparaver

Paraver				
File Hints Help				_
	Preferences			
Workspaces Tu	Global Timelin	e Histogram Color Workspaces External applications		
None 👷 <u>A</u> t	Trace			
Window browser	Fill State gap	s with IDLE State		*
	View full pat	in trace selector		
	Maximum land			
	Maximum load	able trace size (MB) (500		
	Default direct	ories		
	Traces	/home/emercada	Browse	
				lalog
	CFGs	/home/emercada/soft/wxparaver/latest/cfgs	Browse	ny of the
	Filters XML	/home/emercada/soft/wxparaver/latest/share/filters-config	Browse	d and
		,		
	lutorials root	/home/emercada/soft/wxparaver/tutorials	Browse	
	Tmp dir	/home/emercada	Browse	ne " UK "
Files & Window Prope	·			
<i>w</i>	Behaviour —			
	□ Allow only or	e running instance		
▶ ■4.10.4	Automatically	save session every 0 🗄 minutes		als is
	Show dialog	for crashed auto-saved sessions on startup		
▶ □ DIN				•
▷ Crgs	Show help co	incents on a browser		Close
	Disable time	ine zoom with mouse wheel		
r ∎ lib64	1			
			<u>K</u> Cancel <u>C</u> K	
	-			
Paraver files	•			
🛚 🗹 Automatic I	R Force Redraw			



Install Paraver

Start Paraver

laptop\$ paraver/bin/wxparaver

Paraver	
File Hints Help	
🛯 📾 🗐 🔹 <u>H</u> elp Contents	T
Workspaces <u>T</u> utorials	-
None 😚 About	
Window browser	
All Iraces	•
Files & Window Properties	-
Files & Window Properties	
Files & Window Properties	
Files & Window Properties	•
Files & Window Properties	
Files & Window Properties	
Files & Window Properties Files & Window Properties + 0.4.10.4 + 0.4.10.5 + 0.5 + 0.5 + 0.5 + 0	-
Files & Window Properties	
Files & Window Properties Files & Window Properties + 4.10.4 + 4.10.5 + bin + cfgs + include + bib64 + bib	
Files & Window Properties	▲ ↓ ▼







Barcelona Supercomputing Center Centro Nacional de Supercomputación



Paraver Hands-On

Judit Giménez, Lau Mercadal, Germán Llort



2024-09-05

PATC: Performance Analysis and Tools

First steps of analysis

• Copy the trace to your laptop

laptop\$ scp <USER>@hpc:lulesh2.0_s65_27p.* ./

• Load the trace with Paraver

Click on File → Load Trace → Browse to ... "lulesh2.0_s65_27p.prv"





First steps of analysis

- Follow Tutorial #3
 - Introduction to Paraver and Dimemas methodology





Measure the parallel efficiency

- Click on "mpi_stats.cfg"
 - Check the Average for the column labeled "Outside MPI"

utorials									
The first question to answer when analyzing a parallel code run?". The efficiency of a parallel program can be defined by parallelization efficiency and the efficiency obtained in the e regions. These two metrics would be the first checks on the • To measure the parallel efficiency load the configured ofge/mpi/mpi_stats.cfg Th s configuration pops every thread spends in every Mi I call. Look at the glo the outside mpi commit. Chur Average represents the	is "how efficie ased on two a xecution of the proposed m ration file up a table wit bal statistics application p	ent does it spects: the e serial ethodology. MPI call profile @ ICE ID 3D Q @) lulesh2.0-int	el_27p_bind. ☆ Σ ⅔ ⊾	.prv Default	• *	_	_	
efficiency, entry Avg/Max represents the global load b	alance and e					_			
represents the communication efficiency. If any of the	se values are	THREAD 1.17.1	89.43 %	0.13 %	0.10 %	0.10 %	0.20 %	0.00 %	0.63 %
control window to identify the phases and iterations o	f the code.	THREAD 1.18.1	84.35 %	0.09 %	0.07%	0.11 %	1.48 %	0.02 %	0.00 %
		THREAD 1.19.1	91.36 %	0.06 %	0.05 %	0.11 %	0.45 %	0.02 %	0.00 %
 To measure the computation time distribution loa 	d the configur	THREAD 1.20.1	84.62 %	0.09 %	0.08 %	0.10 %	0.84 %	0.01 %	0.00 %
<u>cfgs/general/2dh_usefulduration.cfg</u> This co	nfiguration po	THREAD 1.21.1	83.96 %	0.07 %	0.04 %	0.26 %	0.51 %	0.02 %	1.05 %
histogram of the duration for the computation regions are delimited by the exit from an MPI call and the entr	. The compute	THREAD 1.22.1	85.91 %	0.09 %	0.06 %	0.14 %	0.63 %	0.02 %	0.00 %
histogram does not show vertical lines, it indicates the	computation	THREAD 1.23.1	94.69 %	0.14 %	0.09 %	0.10 %	0.20 %	0.01 %	0.00 %
not balanced. Open the control window to look at the	time distribut	THREAD 1.24.1	88.51 %	0.10 %	0.06 %	0.10 %	1.31 %	0.01 %	0.00 %
correlate both views.		THREAD 1.25.1	95.19 %	0.07 %	0.04 %	0.23 %	0.72 %	0.02 %	0.00 %
		THREAD 1.26.1	90.05 %	0.10 %	0.06 %	0.10 %	0.75 %	0.01 %	0.00 %
	tions of a	THREAD 1.27.1	88.14 %	0.08 %	0.04 %	0.29 %	0.87 %	0.02 %	0.00 %
Parallel efficiency	the compu								
	1 an MP1 ca	Tota	2,424.30 %	2.40 %	2.00 %	4.58 %	17.69 %	0.34 %	4.14 %
to are nexe can. If are instogram doesn't show veraca	lines, it indica	Avera	89.79 %	0.09 %	0.07 %	0.17 %	0.66 %	0.01 %	0.15 %
.	pen die cond	Maxim	98.75 %	0.17 %	0.15 %	0.53 %	1.48 %	0.02 %	1.05 %
Comm efficiency		Minimum	83.90 %	0.04 %	0.04 %	0.10 %	0.20 %	0.00 %	0.00 %
	the IPC time	StDev	3.96 %	0.03 %	0.03 %	0.09 %	0.32 %	0.01 %	0.31 %
		Ava/M	0.91	0.51	0.48	0.32	0.44	0.73	0.15
Load balance							i		



Focus on the iterative part

MPI call profile @ lulesh2.0-intel_27p_bind.prv

MPI call @	lulesh2.	.0-intel	_27p_bind.p	rv				
MPI call @	lulesh2.	.0-intel	_27p_bind.p					
	θ us							3,874,011 us

Click on Open Control Window

ic) 3D 🔍 🎕	н н н	Ϋ Σ ½	Default	•			
THREAD 1.17.1	89.43 %	0.13 %	0.10 %	0.10 %	0.20 %	0.00 %	0.63 %
HREAD 1.18.1	84.35 %	0.09 %	0.07 %	0.11 %	1.48 %	0.02 %	0.00 %
HREAD 1.19.1	91.36 %	0.06 %	0.05 %	0.11 %	0.45 %	0.02 %	0.00 %
HREAD 1.20.1	84.62 %	0.09 %	0.08 %	0.10 %	0.84 %	0.01 %	0.00 %
HREAD 1.21.1	83.96 %	0.07 %	0.04 %	0.26 %	0.51 %	0.02 %	1.05 %
HREAD 1.22.1	85.91 %	0.09 %	0.06 %	0.14 %	0.63 %	0.02 %	0.00 %
HREAD 1.23.1	94.69 %	0.14 %	0.09 %	0.10 %	0.20 %	0.01 %	0.00 %
HREAD 1.24.1	88.51 %	0.10 %	0.06 %	0.10 %	1.31 %	0.01 %	0.00 %
HREAD 1.25.1	95.19 %	0.07 %	0.04 %	0.23 %	0.72 %	0.02 %	0.00 %
HREAD 1.26.1	90.05 %	0.10 %	0.06 %	0.10 %	0.75 %	0.01 %	0.00 %
HREAD 1.27.1	88.14 %	0.08 %	0.04 %	0.29 %	0.87 %	0.02 %	0.00 %
Total	2,424.30 %	2.40 %	2.00 %	4.58 %	17.69 %	0.34 %	4.14 %
Average	89.79 %	0.09 %	0.07 %	0.17 %	0.66 %	0.01 %	0.15 %
Maximum	98.75 %	0.17 %	0.15 %	0.53 %	1.48 %	0.02 %	1.05 %
Minimum	83.96 %	0.04 %	0.04 %	0.10 %	0.20 %	0.00 %	0.00 %
StDev	3.96 %	0.03 %	0.03 %	0.09 %	0.32 %	0.01 %	0.31 %
Avg/Max	0.91	0.51	0.48	0.32	0.44	0.73	0.15



Focus on the iterative part



Drag & drop on this area to zoom on the iterative region

PI call profile @ lulesh2.0-intel_27p_bind.prv										
e id 3d 🔾 🎕	н н н	🗙 Σ ዥ 🕨	Default	• \$						
HREAD 1.17.1	89.43 %	0.13 %	0.10 %	0.10 %	0.20 %	0.00 %	0.63 %			
HREAD 1.18.1	84.35 %	0.09 %	0.07 %	0.11 %	1.48 %	0.02 %	0.00 %			
HREAD 1.19.1	91.36 %	0.06 %	0.05 %	0.11 %	0.45 %	0.02 %	0.00 %			
HREAD 1.20.1	84.62 %	0.09 %	0.08 %	0.10 %	0.84 %	0.01 %	0.00 %			
HREAD 1.21.1	83.96 %	0.07 %	0.04 %	0.26 %	0.51 %	0.02 %	1.05 %			
HREAD 1.22.1	85.91 %	0.09 %	0.06 %	0.14 %	0.63 %	0.02 %	0.00 %			
HREAD 1.23.1	94.69 %	0.14 %	0.09 %	0.10 %	0.20 %	0.01 %	0.00 %			
HREAD 1.24.1	88.51 %	0.10 %	0.06 %	0.10 %	1.31 %	0.01 %	0.00 %			
HREAD 1.25.1	95.19 %	0.07 %	0.04 %	0.23 %	0.72 %	0.02 %	0.00 %			
HREAD 1.26.1	90.05 %	0.10 %	0.06 %	0.10 %	0.75 %	0.01 %	0.00 %			
HREAD 1.27.1	88.14 %	0.08 %	0.04 %	0.29 %	0.87 %	0.02 %	0.00 %			
Total	2,424.30 %	2.40 %	2.00 %	4.58 %	17.69 %	0.34 %	4.14 %			
Average	89.79 %	0.09 %	0.07 %	0.17 %	0.66 %	0.01 %	0.15 %			
Maximum	98.75 %	0.17 %	0.15 %	0.53 %	1.48 %	0.02 %	1.05 %			
Minimum	83.96 %	0.04 %	0.04 %	0.10 %	0.20 %	0.00 %	0.00 %			
StDev	3.96 %	0.03 %	0.03 %	0.09 %	0.32 %	0.01 %	0.31 %			
Avg/Max	0.91	0.51	0.48	0.32	0.44	0.73	0.15			



Recalculate efficiency of iterative region



Right click	
→ Сору	

5 ID 30 🔍 🍭		* Σ ½	Default	- 2			
HREAD 1.17.1	89.43 %	0.13 %	0.10 %	0.10 %	0.20 %	0.00 %	0.63 %
HREAD 1.18.1	84.35 %	0.09 %	0.07 %	0.11 %	1.48 %	0.02 %	0.00 %
HREAD 1.19.1	91.36 %	0.06 %	0.05 %	0.11 %	0.45 %	0.02 %	0.00 %
HREAD 1.20.1	84.62 %	0.09 %	0.08 %	0.10 %	0.84 %	0.01 %	0.00 %
HREAD 1.21.1	83.96 %	0.07 %	0.04 %	0.26 %	0.51 %	0.02 %	1.05 %
HREAD 1.22.1	85.91 %	0.09 %	0.06 %	0.14 %	0.63 %	0.02 %	0.00 %
HREAD 1.23.1	94.69 %	0.14 %	0.09 %	0.10 %	0.20 %	0.01 %	0.00 %
HREAD 1.24.1	88.51 %	0.10 %	0.06 %	0.10 %	1.31 %	0.01 %	0.00 %
HREAD 1.25.1	95.19 %	0.07 %	0.04 %	0.23 %	0.72 %	0.02 %	0.00 %
HREAD 1.26.1	90.05 %	0.10 %	0.06 %	0.10 %	0.75 %	0.01 %	0.00 %
HREAD 1.27.1	88.14 %	0.08 %	0.04 %	0.29 %	0.87 %	0.02 %	0.00 %
Total	2,424.30 %	2.40 %	2.00 %	4.58 %	17.69 %	0.34 %	4.14 %
Average	89.79 %	0.09 %	0.07 %	0.17 %	0.66 %	0.01 %	0.15 %
Maximum	98.75 %	0.17 %	0.15 %	0.53 %	1.48 %	0.02 %	1.05 %
Minimum	83.96 %	0.04 %	0.04 %	0.10 %	0.20 %	0.00 %	0.00 %
StDev	3.96 %	0.03 %	0.03 %	0.09 %	0.32 %	0.01 %	0.31 %
Avg/Max	0.91	0.51	0.48	0.32	0.44	0.73	0.15



Recalculate efficiency of iterative region

call @	<pre>lulesh2.0-intel_27p_bind.prv</pre>								
					nallılır. İlmə.	i			
	547,635 us	MDT coll profile	@ luloch2 Q int	al 27n hind	3,5	04,864 us			_
			H H H	X Σ ½	Default	- #			
		THREAD 1.17.1	89.03 %	0.12 %	0.09 %	0.10 %	0.20 %	10.41 %	
		THREAD 1.18.1	83.71 %	0.09 %	0.06 %	0.06 %	1.58 %	14.44 %	
		THREAD 1.19.1	91.09 %	0.06 %	0.04 %	0.09 %	0.46 %	8.21 %	
	Dight click	THREAD 1.20.1	84.05 %	0.09 %	0.08 %	0.06 %	0.88 %	14.79 %	
	Right Click	THREAD 1.21.1	83.33 %	0.07 %	0.04 %	0.25 %	0.52 %	15.73 %	
		THREAD 1.22.1	85.40 %	0.09 %	0.06 %	0.10 %	0.65 %	13.66 %	
	\rightarrow Paste \rightarrow	THREAD 1.23.1	94.58 %	0.14 %	0.09 %	0.09 %	0.19 %	4.87 %	
		THREAD 1.24.1	88.12 %	0.10 %	0.06 %	0.06 %	1.41 %	10.20 %	
	Time	THREAD 1 25.1	95.10 %	0.07 %	0.04 %	0.23 %	0.76 %	3.76 %	
	Time	THREAD 1.26.1	<u>۹۹.73 %</u>	0.10 %	0.05 %	0.05 %	0.80 %	9.23 %	
		THREAD 1.27.1	87.66 %	0.7%	0.04 %	0.31 %	0.92 %	10.95 %	
		Total	2,414.73 %	2.29 %	1.96 %	4.09 %	18.55 %	257.03 %	
		Average	89.43 %	0.08 %	0.07 %	0.15 %	0.69 %	9.52 %	
		Maximum	98.76 %	0.17 %	0.15 %	0.55 %	1.58 %	16.08 %	
		Minimum	83.33 %	0.04 %	0.04 %	0.05 %	0.19 %	0.01 %	
		StDev	4.15 %	0.03 %	0.03 %	0.11 %	0.35 %	4.09 %	
		Avg/Max	0.91	0.51	0.48	0.28	0.44	0.59	
									_



0.05 % 0.05 % 0.05 % 0.05 % 0.05 % 0.05 % 0.05 % 0.05 % 0.05 %

1.35 % 0.05 % 0.06 % 0.05 % 0.00 % 0.87

Efficiency of iterative region

		MPI call profile	@ lulesh2.0-int	tel_27p_bind	.prv					
IE ID 30 Q € I H H H II Y Σ ½ ► Default ▼ ∅										
		THREAD 1.17.1	89.03 %	0.12 %	0.09 %	0.10 %	0.20 %	10.41 %	0.05 %	
		THREAD 1.18.1	83.71 %	0.09 %	0.06 %	0.06 %	1.58 %	14.44 %	0.05 %	
		THREAD 1.19.1	91.09 %	0.06 %	0.04 %	0.09 %	0.46 %	8.21 %	0.05 %	
		THREAD 1.20.1	84.05 %	0.09 %	0.08 %	0.06 %	0.88 %	14.79 %	0.05 %	
		THREAD 1.21.1	83.33 %	0.07 %	0.04 %	0.25 %	0.52 %	15.73 %	0.05 %	
		THREAD 1.22.1	85.40 %	0.09 %	0.06 %	0.10 %	0.65 %	13.66 %	0.05 %	
		THREAD 1.23.1	94.58 %	0.14 %	0.09 %	0.09 %	0.19 %	4.87 %	0.05 %	
	_	THREAD 1.24.1	88.12 %	0.10 %	0.06 %	0.06 %	1.41 %	10.20 %	0.05 %	
		THREAD 1.25.1	95.10 %	0.07 %	0.04 %	0.23 %	0.76 %	3.76 %	0.05 %	
Parallel efficiency		THREAD 1.26.1	89.73 %	0.10 %	0.05 %	0.05 %	0.80 %	9.23 %	0.05 %	
		THREAD 1.27.1	87.66 %	0.07 %	0.04 %	0.31 %	0.92 %	10.95 %	0.06 %	
Comm officionau		Το. '	2,414.73 %	2.29 %	1.96 %	4.09 %	18.55 %	257.03 %	1.35 %	
comm endency		Averare	89.43 %	0.08 %	0.07 %	0.15 %	0.69 %	9.52 %	0.05 %	
		Maxir	98.76 %	0.17 %	0.15 %	0.55 %	1.58 %	16.08 %	0.06 %	
		Minimum	83.33 %	0.04 %	0.04 %	0.05 %	0.19 %	0.01 %	0.05 %	
Load balance		StDev	4.15 %	0.03 %	0.03 %	0.11 %	0.35 %	4.09 %	0.00 %	
		Avg/I**	0.91	0.51	0.48	0.28	0.44	0.59	0.87	



Computation time distribution

• Click on "2dh_usefulduration.cfg" (2nd link) → Shows time computing





Focus on the iterative part

Click on "2dh_usefulduration.cfg" (2nd link) → Shows time computing





Computation time distribution

Click on "2dh_usefulduration.cfg" (2nd link) → Shows time computing





Computation time distribution

Click on "2dh_usefulduration.cfg" (2nd link) → Shows time computing





Computation load distribution

• Click on "2dh_useful_instructions.cfg" (3rd link) → Shows **amount of work**





Correlate two histograms

• Clear correlation between the **amount of work** and the **time computing**





Go from the table to the timeline





Go from the table to the timeline





• Slow & Fast at the same time? \rightarrow Imbalance



Zoom into 1 of the iterations (by drag-and-dropping)



39



Hints → Call stack references → Caller function







• Hints \rightarrow Call stack references \rightarrow Caller function





Save CFG's (method 1)



Copy Paste	Ctrl+C	
Clone		
Undo Zoom	Ctrl+U	
Redo Zoom Fit Time Scale	Ctrl+R	
Fit Semantic Scale Fit Objects Select Objects	Þ	
View Paint As	> >	
Drawmode Pixel Size	¢	
Object Labels	Þ	
Object Axis	Þ	
Run	Þ	
Synchronize Remove all sync		
Save	•	Configuration
Info Panel		lmage Image Legend Text



Save CFG's (method 2)



Centro Nacional de Supercomputación

CFG's distribution

• Paraver comes with many more included CFG's

😣 🗏 🗉 🛛 Paraver		
File Hints Help		
Load Trace	Ctrl+O	
Previous Traces	•	
Unload Traces		
Load Configuration		
Previous Configurations	•	/home/gllort/Apps/wxparaver/latest/cfgs/General/views/useful_duration.cfg
Save Configuration		/home/gllort/Apps/wxparaver/4.6.2/cfgs/counters_PAPI/performance/2dh_cycles_per_us.cfg
Load Session	Ctrl+L	/home/gllort/Apps/wxparaver/4.6.2/cfgs/mpi/analysis/mpi_stats.cfg
Save Session	Ctrl+S	/home/gllort/Apps/wxparaver/latest-tutorials/3.Introduction_to_Paraver_and_Dimemas_methodology/cfgs/papi/2dh_useful_instructions.cfg
Preferences		/home/gllort/Apps/wxparaver/latest/cfgs/counters_PAPI/performance/cycles_per_us.cfg
Quit	chaluo	/home/gllort/Apps/wxparaver/4.6.2/cfgs/clustering/2dp_clusters.cfg
Quic	Ctrt+Q	/home/gllort/Apps/wxparaver/latest-tutorials/3.Introduction_to_Paraver_and_Dimemas_methodology/cfgs/general/2dh_usefulduration.cfg
		/home/gllort/Apps/wxparaver/4.6.2/crgs/counters_PAPI/performance/2dn_userulduration.crg
		/home/gllort/Apps/wxparaver/4.6.2/cfgs/counters_PAPI/performance/2dh_useful_instructions.cfg
		/home/gllort/Apps/wxparaver/4.6.2/cfgs/General/sanity_checks/flushing.cfg
		/home/gllort/Apps/wxparaver/4.6.2/cfgs/counters_PAPI/performance/IPC.cfg
		/home/gllort/Apps/wxparaver/latest/cfgs/General/views/executing_cpu.cfg
Fles S. Window Properties		/home/gllort/Apps/wxparaver/4.6.2/cfgs/clustering/3dh_duration_cid.cfg
		/home/gllort/Apps/wxparaver/latest/cfgs/clustering/clusterID_window.cfg
		/home/gllort/Apps/wxparaver/latest-tutorials/3.Introduction_to_Paraver_and_Dimemas_methodology/cfgs/mpi/mpi_stats.cfg
tracking		/home/gllort/Apps/wxparaver/latest/cfgs/General/views/user_functions.cfg
		/home/gllort/Apps/wxparaver/cfgs/memory_location.cfg
► = 4.6.2		/home/gllort/Apps/wxparaver/cfgs/store_samples.cfg
Fight Star		/home/gllort/Apps/wxparaver/cfgs/load_samples.cfg
▼		/home/gllort/Apps/wxparaver/cfgs/memkind_partition.cfg
▶ ■ bin		
E Durch mode		
 buist_mode clustering 		
r 📕 clustering		



Paraver files

Counters PAPI

*

Hints: a good place to start!

• Paraver suggests CFG's based on the contents of the trace





• Paraver is like an equation editor, and CFG's like a formula

• Filter

• The data I'm interested on seeing (Example: Instructions)

Files &	Files & Window Properties			40000003 Flushing Traces					
	Click of the second sec			Salact	8 Tracing mode:				
					B Executing CPU				
🖯 Fil	ter		from th	ne list	6 Process IDentifier				
\pm	Communication	15			7 Parent Process IDentifier				
	Events			400000)38 fork() depth				
	Event type	=; 42000050		40000	050 RAW clock() value from system				
	Function	=		400001	133 CPU-Event sampling interval				
	Types	42000050		410000	000 Object addresses for task 1.1				
	Type/Value Op	And		419999	999 Active bardware counter set				
Ξ	Event value	All;		420000	2000000 PARI 11 DCM [11D cacho missos]				
	Function	All		420000	02 PAPI_L2_DCM [L2D cache misses]				
	Values			420000					
				420000	008 PAPI_L3_TCM [L3 cache misses]				
				420000	050 PAPI_TOT_INS [Instr completed]				
Ba	rcelona			420000	059 PAPI_TOT_CYC [Total cycles]				



- Semantic
 - How to paint the filtered data
 - Next Evt Val?



iles & Window Properties							
Semantic Semantic							
Top Con	npose 1	As Is					
Top Con	npose 2	As Is					
Compos	Compose Thread As Is						
Thread		Next Evt Val					

Each measurement counts how many Instructions were executed in preceding region

- Instructions count is found at the end of the region
- Next Evt Val = Color is determined by the next event value





 Derived Instructions @ lulesh_ Instructions 	27p.prv	Create Deriv	ed Timeline Wind 😣
• Cycles @ lulesh_27g THREAD 1.1.1 THREAD 1.19.1 THREAD 1.19.1 THREAD 1.27.1 0 us 465,691,125 -	772,606,787 7,359.168 ut	1 As Is	•
 Instructions / Cycles = IPC 			
Paraver – 😣 File Hints Help		2 As Is	•
Workspaces	I	Factor	1,000000
Useful+MPI+PAPI countersSelect InstructionsWindow browserand drag-and-droplulesh_27p.prvand drag-and-drop		Timeline	Instructions
Instructions On Cycles		Operation	divide 👻
	Operation:	Timeline	Cycles
	divide	Factor	1,000000
		Sv	vap Windows
BSC Barcelona Supercomputing Center			Cancel OK

.359.168.443

- Derived
 - Instructions ۲
 - Cycles
 - Instructions / Cycles = IPC ٠

HREAD 1.1.1 HREAD 1.10.

HREAD 1.19.

HREAD 1.1.1

THREAD 1.10.1 HREAD 1.19



Instructions @ lulesh_27p.prv

cycles @ lulesh_27p.prv



Build any formula with the data available and save a CFG!



Main views of Paraver (I)

Timeline

Processes (and threads)



Main views of Paraver (II)

• Table (Profile)

Categories (e.g. MPI calls)

		MPI call profile @ lulesh2.0_27p.prv 🗆 🧍																				
	IC.	D	30	Q,			н	+4		*	Σ	₽∕E		D	efault	-	41					
				Out	side	MPI	MPI_	Isen	d M	PI_Ire	ecv	MPI.	Wai	it	MPI_W	aitall	MPI_I	Barrier	MPI	Reduc	e MP	
	тни	READ	1.1.1		99.0	04 %		0.05	%	0.0	6%	0).35 °	%	C	0.30 %		0.03 %	•	0.00	%	
	THE	READ	1.2.1		97.3	37 %		0.07	%	0.0	8 %	0).20 °	%	C	0.82 %		0.03 %	•	0.00	%	
	THE	READ	1.3.1		93.7	79 %		0.05 (%	0.0	5 %	C).22 °	%	C).52 %		0.03 %		0.02	%	
	THE	READ	1.4.1		93.9	93 %		0.07	%	0.0	8 %	().17 ⁽	%	C	0.61 %		0.03 %	•	0.00	%	
	THE	READ	1.5.1		93.7	75 %		0.11 '	%	0.1	1 %	0	D.38 °	%	C	0.19 %		0.01 %	•	0.00	%	
	THE	READ	1.6.1		91.6	54 %		0.08	%	0.0	8 %	0).10 ⁽	%	C	0.74 %		0.02 %	•	0.00	%	
•	THE	READ	1.7.1		91.2	24 %		0.06	%	0.0	5 %	0).16 ⁽	%	C	0.42 %		0.03 %	•	0.11	%	
	THE	READ	1.8.1		91.9	93 %		0.08	%	0.0	8 %	0	0.16	%	C).76 %		0.03 %		0.00	%	
	THE	READ	1.9.1		91.2	20 %		0.06	%	0.0	5 %	().14 °	%	C).59 %		0.02 %	•	0.50	%	
	THR	EAD	1.10.1	1	90.4	17 %		0.08	%	0.0	7 %	0).33 °	%	C	0.37 %		0.03 %	•	0.00	%	
	THR	EAD	1.11.1	1	89.1	19 %		0.12	%	0.1	1 %	0).35 °	%	C	0.30 %		0.01 %	•	0.70	%	
	THR	EAD	1.12.1	1	95.8	30 %		0.09 (%	0.0	7 %	0).19 ^o	%	C	0.83 %		0.03 %	•	0.00	%	
	THR	EAD	1.13.1	1	96.0	04 %		0.12	%	0.1	0 %	0).41 ^o	%	C).33 %		0.01 %	•	0.00	%	
	THR	EAD	1.14.1	1	94.6	51 %		0.18	%	0.1	5 %	C).15 °	%	C	0.35 %		0.00 %	•	0.00	%	
-	THR	EAD	1.15.1	1	93.2	28 %		0.13	%	0.1	0 %	0).10 ⁽	%	1	.21 %		0.01 %	•	0.00	%	
	THR	EAD	1.16.1	1	91.6	57 %		0.09	%	0.0	7 %	0	0.26	%	2	2.01 %		0.03 %	•	0.00	%	
	THR	EAD	1.17.1	1	93.2	28 %		0.13	%	0.1	0 %).11 ⁽	%	1	.06 %		0.01 %		0.45	%	
	THR	EAD	1.18.1	1	89.5	56 %		0.09	%	0.0	7 %	0	D.16 °	%	1	.72 %		0.03 %	•	0.00	%	
	THR	EAD	1.19.1	1	94.0	06 %		0.06	%	0.0	4 %	0).13 ⁽	%	C	0.47 %		0.03 %		0.00	%	
	THR	EAD	1.20.1	1	89.3	39 %		0.10	%	0.0	б%	0).25 °	%	1	.05 %		0.03 %	•	0.00	%	
	THR	EAD	1.21.1	1	89.6	52 %		0.07	%	0.0	4 %	0).22 °	%	C	0.30 %		0.03 %	•	0.90	%	
	THR	EAD	1.22.1	1	88.0	08 %		0.09	%	0.0	6 %	0	0.26 9	%	2	2.02 %		0.03 %	•	0.00	%	
	THR	EAD	1.23.1	1	98.1	19 %		0.14	%	0.1	0 %	0).16 °	%	C	0.62 %		0.01 %		0.00	%	
	THR	EAD	1.24.1	1	94.1	10 %		0.10	%	0.0	6 %).12 °	%	1	.24 %		0.02 %	•	0.00	%	
	THR	EAD	1.25.1	1	96.0	05 %		0.07	%	0.0	4 %	(0.29 9	%	C	0.26 %		0.02 %		0.00	%	
	THR	EAD	1.26.1	1	93.1	10 %		0.10	%	0.0	6%	().13 ⁽	%	1	.13 %		0.03 %		0.00	%	
	THR	EAD	1.27.1	1	94.2	24 %		0.08 9	%	0.0	4 %	(0.18 °	%	C).39 %		0.02 %		0.00	%	
		Tota	al	2	,514.6	52 %		2.44	%	1.9	9 %	5	5.69	%	20	0.63 %		0.60 %	•	2.72	%	
	-	Avera	ige		93.1	13 %		0.09	%	0.0	7%	0	0.21	%	C	0.76 %		0.02 %	•	0.10	%	
	N	laxim	um		99.0	04 %		0.18	%	0.1	5 %	0	0.41	%	2	2.02 %		0.03 %	•	0.90	%	
	N	linim	um		88.0	08 %		0.05	%	0.0	4%	0	0.10	%	C	0.19 %		0.00 %	•	0.00	%	
		StDe	ev		2.7	79 %		0.03	%	0.0	3 %	0	0.09	%	C	0.51 %		0.01 %	•	0.24	%	
	4	vg/N	/lax			0.94		0.5	0	C	0.49		0.5	52		0.38		0.72	2	0.1	11	

The table can display a variety of statistics (e.g. % of time, # of calls, etc.) with gradient coloring showing from low values to high values

Summary



Main views of Paraver (II

Histograms

Having data in the middle of the histogram indicates that the program executes medium-number of а instructions.

Being the data scattered (to the left and right), indicates that different processes do different amounts of work (left \rightarrow less; right \rightarrow more).



Having data in the right of the X-axis indicates that the program executed a very high number of instructions.

The vertical line indicates that all processes did the same amount of work, as there are no variations along the Y-axis.

Gradient from green to blue indicates low to high values of selected statistic (default is **Time**) :



The more to the left \rightarrow doing less instructions \rightarrow greener \rightarrow low % of exec time Data on the right \rightarrow doing more instructions \rightarrow bluer \rightarrow high % of exec time



Statistic: % of time (Color)





Statistic: # of bursts (Color)



Proc 1: 1M	1M	5M	1M	1M	5M
Proc 2: 1M	1M	5M	1M	1M	5M
Proc 3: 1M	1M	5M	1M	1M	5M
Proc 4: 1M	1M	5M	1M	1M	5M

	0-1 M	1-2 M	2-3M	3-4M	4-5M	
Proc 1:	(4)				(2)	
Proc 2:	(4)				(2)	Metric:
Proc 3:	(4)				(2)	(Columns)
Proc 4:	(4)				(2)	

Statistic: # of bursts (Color)



Proc 1:	1M 1M		1M	1M		5M				
Proc 2: [1M 1M		5M	1M	1M 1M			5M		
Proc 3:	2M	2M	2M 2M		21	N	2M	2M		
Proc 4: [1M 1M		5M	1M	1M		5M			

	0-1 M	1-2 M	2-3M	3-4M	4-5M	
Proc 1:	(4)				(2)	
Proc 2:	(4)				(2)	Metric:
Proc 3:			(7)		(2)	(Columns)
Proc 4:	(4)				(2)	

Statistic: # of bursts (Color)





Barcelona Supercomputing Center Centro Nacional de Supercomputación



Getting traces for different scales

Judit Giménez, Lau Mercadal, Germán Llort



PATC: Performance Analysis and Tools

2024-09-05

Let's get more traces!

• What happens if we bind processes to cores?

```
#!/bin/bash
#SBATCH --job-name=lulesh2.0 s65 27p bound
#SBATCH --output=%x %j.out
#SBATCH --error=%x %j.err
#SBATCH --ntasks=27
#SBATCH --time=00:10:00
module load extrae
export TRACE_NAME=lulesh2.0_s65_27p_bound.prv
srun --cpu-bind=cores -m cyclic:cyclic \
     ./trace.sh lulesh2.0 -i 10 -s 65 -p
```



Let's get more traces!

• And if we run with 8 and 64 processes?

```
#!/bin/bash
#SBATCH --job-name=lulesh2.0s65_8p_bound
#SBATCH --output=%x %j.out
#SBATCH --error=%x %j.err
#SBATCH --ntasks=8
#SBATCH --time=00:10:00
module load extrae
export TRACE_NAME=lulesh2.0_s65_8p_bound.prv
srun --cpu bind=cores -m cyclic:cyclic \
     ./trace.sh ./lulesh2.0 -i 10 -s 65 -p
```



And cut them to select the FoA







Barcelona Supercomputing Center Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Getting OpenMP traces

Judit Giménez, Lau Mercadal, Germán Llort



2024-09-05

PATC: Performance Analysis and Tools

Let's get MPI+OpenMP traces!

• What changes do I need in my jobscript?

```
#!/bin/bash
#SBATCH --job-name=lulesh2.0 s65 8p bound
#SBATCH --output=%x %j.out
#SBATCH --error=%x %j.err
#SBATCH --ntasks=8
#SBATCH --cpus-per-task=8
#SBATCH --time=00:10:00
export OMP NUM THREADS=8
export TRACE NAME=lulesh2.0 s65 8p8omp bound.prv
srun --cpu-bind=cores -m cyclic:cyclic \
     ./trace.sh lulesh2.0-omp -i 10 -s 65 -p
```



Let's get MPI+OpenMP traces!

• What changes do I need in my XML config file?

```
<mpi enabled="yes">
    <counters enabled="yes" />
    </mpi>
<openmp enabled="yes">
    <locks enabled="no" />
    <counters enabled="yes" />
</openmp>
<pthread enabled="no" />
    <locks enabled="no" />
    <counters enabled="yes" />
</pthread>
```



Let's get MPI+OpenMP traces!

• What changes do I need in my `trace.sh` file?

```
#!/bin/bash
# Configure Extrae
export EXTRAE HOME=<installation base path>
export EXTRAE CONFIG FILE=./extrae.xml
# Load the tracing library (choose C/Fortran)
export LD PRELOAD=$EXTRAE HOME/lib/libompitrace.so
#export
LD PRELOAD=$EXTRAE HOME/lib/libmpitracef.so
# Run the program
$*
```



OpenMP Workspace

• Paraver detects OpenMP and suggests views through Hints

Parave	r			
<u>F</u> ile H	Hints <u>H</u> elp			
i 💽	Useful	•		
Works	MPI			
Usef	OpenMP	\mathbf{F}	Para	allel functions in user code
Windo	PAPI counters	►	Out	lined functions profile
lules	Call stack references	•	Hist	ogram of outlined functions duration
Files & V	Vindow Properties			
2				





Barcelona Supercomputing Center Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Thank you!

Judit Giménez, Lau Mercadal, Germán Llort



2024-09-05

PATC: Performance Analysis and Tools