

Comparing RDataFrame and PROOF Lite

D. Piparo, J. Cervantes (CERN, EP-SFT)

ROOT

Data Analysis Framework

<https://root.cern>

- ▶ Fons produced a comparison between RDF and TSelector/PROOF
 - Based on “h1 analysis” tutorial
 - Very comprehensive and reproducible (thanks Fons!)
 - Full description attached to the agenda
- ▶ This talk shows what we learned from it
- ▶ Partial conclusions can be drawn, but work is still ongoing



Fons' Numbers in a Nutshell

run_selector.C real time: 15 sec

run_selector_proof.C real time: 15 sec (4 cores - speedup factor 1.0)

14 sec (8 cores)

14 sec (12 cores)

14 sec (24 cores)

14 sec (48 cores)

run_dataframe.C real time 56 sec

run_dataframe_imt.C real time 165 sec (4 cores - speedup factor 0.3)

170 sec (8 cores)

167 sec (12 cores)

167 sec (24 cores)

165 sec (48 cores)



Reproduce the Study

- ▶ Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz
- ▶ ROOT compiled in ReleaseWithDebInfo

TSelector	4s
RDF Sequential jitted filters/defines	10s
RDF Sequential explicit lambdas	30s
RDF Sequential explicit lambdas Aclic	4s



Reproduce the Study

- ▶ Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz
- ▶ ROOT compiled in ReleaseWithDebInfo

TSelector	4s	4s
RDF Sequential jitted filters/defines	10s	6s
RDF Sequential explicit lambdas	30s	5s
RDF Sequential explicit lambdas Aclic	4s	4s

Made the header files in the
\$ROOTSYS/include directory
read-only: turn-off cling
nullptr-check



Effect of Nullptr Check

- ▶ Extremely useful
- ▶ Not for free
- ▶ Benchmarked with IgProf:
 - [W/o check report](#)
 - [With check report](#)

Counter: PERF_TICKS

Rank	% total	Counts		Paths		Symbol name
		to / from this	Total	Including child / parent	Total	
	25.33	4.41	4.70	62	73	cling_runtime_internal_throwIfInvalidPointer
[35]	25.33	0.29	4.12	62	62	cling::utils::platform::IsMemoryValid(void const*)
	23.60	4.11	4.11	56	56	__write_nocancel
	0.07	0.01	0.01	2	2	write

[Back to summary](#)



Apples to Apples

Workers	PROOF/TSel	RDF interp	RDF compiled	RDF compiled no jit
1	67s	69s	61s	49s
4	22s	34s	26s	19s
8	16s	22s	17s	16s
12	19s	18s	15s	12s
24	20s	13s	13s	10s



- ▶ RDF compares very well with the PROOFlite MP approach
- ▶ Scaling need attention: why neither PROOFlite nor RDF scale well?
- ▶ Cling's nullptr check has a price which we might reduce