



Benchmarking and optimising large scale parallel workflows

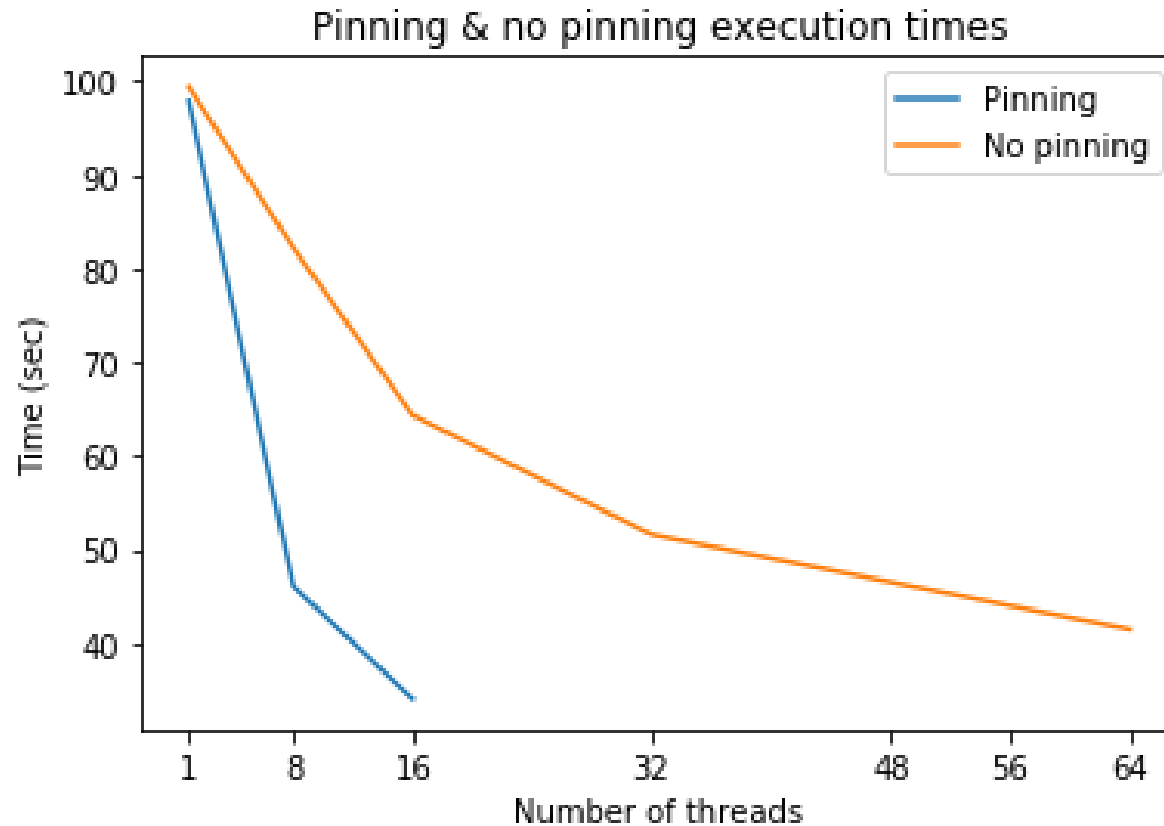
Jesús Perales Hernández

15/08/2019

Overview

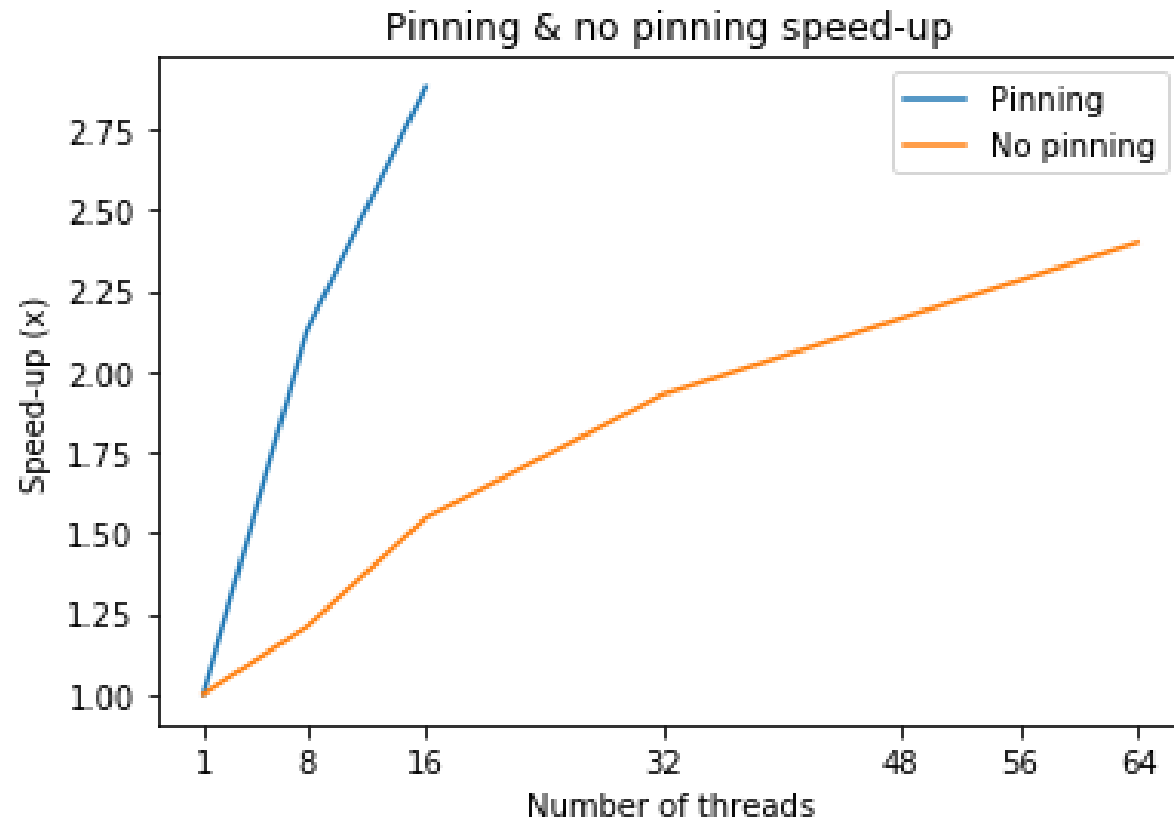
- Nano AOD Higgs Analysis from Stephan Wunsch
- Run on olhswep.cern.ch machine (Xeon(R) E5-2698 @ 2.3GHz) - 32 cores
- Run with ROOT 6.19
- ROOT compiled with -O2 -g -fno-omit-pointer
- Input: 8 files [173 MB to 4.3 GB]
- VTune 2019, perf (visualization with gprof2dot + flamegraph)

Execution times



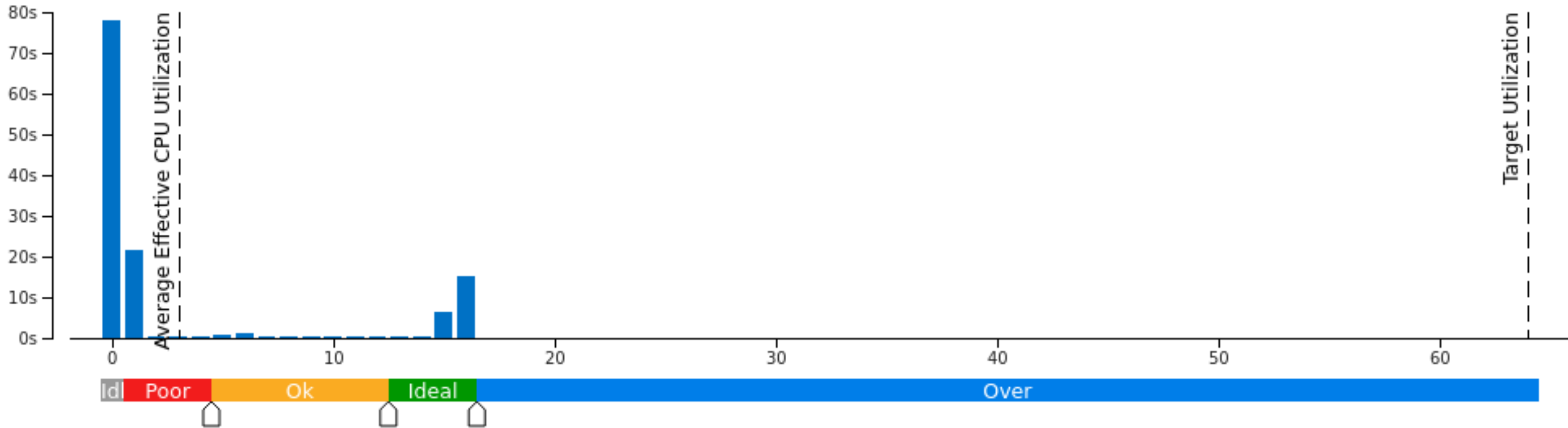
- *Pinning is faster*
- *More threads, lower execution time*

Execution speed-up



- Performance not scaling properly

CPU usage



Some contributions

- There are scalability issues
- The problematic code is identified (and being analysed)
- NUMA effects are important

Impact and future work

- Pull request with the new compiled tutorial
- Improvement in the performance of the RDataFrame
- Improvement in energy & time consumption
- Refactor part of RDataFrame (?)



QUESTIONS?

jesus.perales.hernandez@cern.ch

Jesús Perales-Hernández (LinkedIn)