

WLCG HEP-SCORE Deployment Task Force

Meeting on 15 June 2022 at 15:00 h UTC (teleconference)

Notes

Indico event page: <https://indico.cern.ch/event/1067909/>

Welcome, note-taking, notes from previous meeting, matters arising

Domenico Giordano and Randy Sobie will be co-chairing the Task Force from now on. Minutes of this meeting are taken by Oxana Smirnova. Minutes of the previous meeting are approved (thanks Andrea Valassi for preparing them).

Status of benchmarking campaigns (Gonzalo Menendez Borge)

Gonzalo Menendez Borge reports (see slides) some good progress with testing at sites, though the goal of 10 measurements per **machine** is not quite there yet. Still many yellow and red numbers, which better become green.

Domenico Giordano stresses that it is very important to cover all workloads on all CPUs to get 10 measurements per CPU model in order to propose HEPscore candidates, and encourages sites to expedite testing.

Progress on workloads (Domenico Giordano)

Domenico Giordano focusses on ALICE gen_sim_reco and ATLAS reco_mt on real data workflows (see “Minutes” attached to the contribution in Indico). For the former, some of the problems with zombie processes are understood, site admins are contacted to cross-check. For the latter, some code cross-checking (walltime vs CPUtime usage) needs to be done still.

Walter Lampl confirms that CPUtime was indeed used instead of walltime, and fix is applied. There was the same issue in another workload (atlas_sim_mt), also fixed.

Stefano Piano explains that the breakdown of time reporting into sim, digi and reco (i.e., no gen) is due to the “sim” having simulation and generation together and can not be split, which is enough for the benchmarking purposes.

Randy Sobie asks whether it will be possible to distinguish ATLAS reco_mt on data from others? Domenico answers that these will be in distinct containers.

Survey on HEPscore expectations (Randall Sobie)

Randy Sobie presents an overview of the proposed survey to elicit input on HEPscore expectations (see slides). Results are expected to be presented on the TF meeting on July 6, and initial findings in end-August, to be discussed at the F2F workshop at CERN in September.

Andrea Valassi asks to clarify the meaning of “latest architectures” in Slide 8, and why is it important. Randy explains that tests show that limiting architectures to top-5 removes a lot of scatter in the results, but agrees that it indeed may not matter in terms of the HEPscore definition. Domenico further clarifies that a list of several HEPscore candidates will be prepared, and the one that produces most representative results will be selected. Andrea further asks what would be the procedure of selecting the proper candidate – not defined yet.

Helge Meinhard believes that the question of modern architectures is related to the question of how long would a benchmark be valid. One complete LHC run cycle (5-6 years) should be reasonable, such that work needs not be re-done in mid-run. Newest CPU architectures match this expectation, as the older ones will gradually drop out. Helge also adds that there may be some pressure to include all the workloads because many people invested work into those, but making the test unnecessarily large may not serve the community well, so a subset of workloads would be a good compromise. Other workloads may well be executed as well, not being a part of the benchmark. Regarding running time, Helge believes that 24 hours is perhaps too much, so the chosen subset probably should run within 6 hours. Lastly, he points out that the suggested “Fast benchmark” must be considered very carefully and should be given a proper name that won’t lead to any confusion on its applicability.

Domenico agrees to the comments, particularly that the “Fast benchmark” should not be used as a replacement for purposes of pledges etc., only for some quick testing.

Andrea Valassi adds that on slide 6, numbers in percentage of usage may be more illustrative. Also, an example of a candidate with unequal weights per workflow may also be useful. Randy agrees, and adds that some benchmarks of similar weight can be combined in one (ATLAS and CMS,, for example). Andrea further suggests that weight of simulation should probably be higher than other workflows, for all experiments.

Andrew Melo points out that CMS has a public Wiki with a good breakdown per workflow.

Both Andrea Valassi and Andrew Melo ask whom will the questionnaire be sent to. Randy explains that this version targets the TF members, some of which represent the experiments and presumably can iterate with them. Domenico adds that this survey is a preparation for the workshop, and another survey may well be prepared later, for another audience.

Walter Lampl notes that there’s no such thing as a single ATLAS generator workflow, because there are many, and the chosen one is representative. He further suggests that the walltime for a benchmark test execution should not exceed 6 hours, and probably number of jobs needs to be cut down. For example, if some jobs are strongly correlated, it may be enough to run only one.

Domenico points out that the survey is precisely an opportunity for everybody to submit this kind of comments.

Stefano Piano comments on shortening the workload execution time, pointing out that: Pb-Pb collisions are rather heavy, so it’ll be difficult to make such workflows shorter or parallelise more.

Any other business

Domenico reports on the workshop preparation (see “Minutes” attached to the agenda item): the F2F workshop is scheduled to September 19-20 at CERN, Indico will soon be populated, and contributions will be solicited.

Next meetings

- 1 meeting in July (July 6), and 1 in August, proposed on August 24 (exceptionally). All agreed.
- After summer, the next meeting is on September 7, followed by the Workshop
- Further schedule to be decided after the Workshop.

Annex: Attendance

Present:

Domenico Giordano (CERN; notes)
Walter Lampl (U Arizona)
Helge Meinhard (CERN, chair)
Andrew Melo (Vanderbilt U)
Gonzalo Menendez Borge (CERN)
Gonzalo Merino (PIC)
Stefano Piano (INFN Trieste)
Oxana Smirnova (U Lund)
Randall Sobie (U Victoria)
Andrea Valassi (CERN)
Tony Wong (BNL)
Matthias Schnepf
Yan Xiaofei

Absent:

Miltiadis Alexis (CERN)
Tommaso Boccali (INFN Pisa)
Simone Campana (CERN)
Ian Collier (STFC-RAL)
Alastair Dewhurst (STFC-RAL)
Michel Jouvin (IJCLab)
Andrew McNab (U Manchester)
Bernd Panzer-Steindel (CERN)
Fazhi Qi (IHEP)
Jeff Templon (Nikhef)
Josh Willis (Caltech)

Apologies: