

WLCG HEP-SCORE Deployment Task Force

Meeting on 06 April 2022 at 16:00 h CEST (teleconference)

Notes

Indico event page: <https://indico.cern.ch/event/1104202/>

Welcome, note-taking, notes from previous meeting

The minutes from the previous meeting are approved; thanks to Gonzalo Menendez Borge for having provided them. Nothing to be added to the agenda.

CPU types in ATLAS and CMS jobs and their coverage by benchmarking platforms (Andrea Scialbà)

Andrea Scialbà has been looking into the CMS and ATLAS job logs extracting information on used CPU models. The last three months ATLAS and CMS job records in Elastic Search have been considered to extract the used CPU architectures in the different sites, including HPC sites and excluding BOINC jobs to avoid consumer CPUs. Rome is the largest used CPU architecture followed by Intel architectures: Broadwell, Haswell, Cascade Lake and Skylake. Everything else accounts for a small fraction. For the top 12 CPU architectures, the top two sites in terms of number of job runs have been listed. From such a list we know in which site we can run the benchmark for a specific architecture. In addition, Andrea showed all the architectures tested by the HEPiX Benchmarking Working Group (AMD, Intel and ARM) and he commented also his backup slides that contain the full list of architectures used by ATLAS and CMS jobs, where the correspond architecture is highlighted in green if it has been tested by HEPiX.

Helge commented that considering ALICE and LHCb jobs will not change the picture in any major way.

Tommaso noticed that in the architecture list the Power architectures is missing but CMS has been running on Power 9 on CINECA. Andrea didn't apply any cut on architecture, most likely the monitoring is not reporting the jobs on Power 9. Helge would consider Power architecture with lower priority. Andrea and Domenico confirmed that we have not any build of HEP-SCORE for Power architectures and HS06 needs to be migrated to Power 9.

Helge commented on the scaling for the number of the cores, he reminded that the whole exercise is really to understand the value of a given machine in the ecosystem. That means not only the CPU, but also all the rest of machine.

Domenico remarked that if we have to give priority, he will access to two new CPU models, as Ice Lake and Milan. These architectures are the most interesting, they will probably in the future represent an important part of the procurement. Helge replied that the focus is less on the absolutely newest models, but we need to cover all the potential use cases we have for the benchmarking. We have now the material we need in order to cross check the list of benchmarking platforms. Most of the platforms we had available on Gonzalo's list until very recently were only provided by a single site. Now we can extend such a list.

Analysis of further benchmark results (Randall Sobie)

Randall gave a quick update on some of the EPSpec06 and SPEC2017 results and he focused on workload results. He showed a summary of all the measurements we have on 48 independent systems labeled by its CPU, the number of cores, hyperthreaded or not. If multiple RAM-values, then the highest value is selected. In addition, Andrea provided him with a mapping between the CPU and the architecture. Randall noticed in the plot that the AMD 7302 has a performance of about 50% more than any other CPU. A histogram for each considered system is created for

each benchmark to remove spurious data. Typically if there's data outside of 10% of a measure the mean they throw it out, but much less than 1% of the data are removed. They updated all the HS-64 vs HS-32 plot and SP-INT vs SP-CPP plot. In addition, Randall showed the workload results. There are eight validated workloads of five experiments. They produced tables of the CPU, the architecture, and the benchmarks where 31-38 different systems are considered with 1600-2600 individual measurements for each workload. They made plots of workload results vs HS06 64 bit, HS06 32 bit, SPEC ccp and int, all of them with a linear fit. In addition, they made the residual plots where the fraction of deviation of the points from the fit lines are plotted. Finally, they defined a figure of merit, which is the average deviation from the fit for each of the eight workloads with each of the four benchmarks. Such a plot reflects that there is more scatter in the Juno and the LHCb benchmark compared to the others. The others are in the area of 5 to 10% range, while Juno and LHCb are in the 20% range. The next step is to get a bit more feedback and then decide what we want to put into the HEP-SCORE.

Helge commented that we can pinpoint the major discrepancies to certain CPU models, while a number of other CPU models seems to behave very much as expected. He is looking forward to completing this analysis and include additional workload.

Domenico remarked that Juno and LHCb difference discrepancy can come from the single workload results, while multi-threaded application can perform better. In addition, Juno workload requires more than 2 GB per core. CMS is the one that apparently fits better.

Andrea V. specified that the figure of merit is just an indication that some of these workloads scale better and are more consistent with the considered benchmarks.

About Juno discrepancy Gonzalo asked if we also see a corresponding spread in the workload results, but Randall replied that typically, the spread of each measurement is 1% and often better than 1%.

Short update on benchmarking runs and on workloads (Gonzalo Menéndez Borge and Domenico Giordano)

Gonzalo showed the overall benchmark progress, where there is quite some progress. About the HEP-SCORE scripts only the column D is blank. That is because the script is still work in progress, and we expect to have it probably shortly after the Easter break. In addition, there's been quite some progress with sites that were struggling a bit. He is trying to help the sites so that we can progress as much as possible. About the contributions that we have of SPEC benchmarks, the progress in the last few weeks has lowered, but it is to be expected since a lot of sites have finished. Similar considerations looking at to the individual metrics. The differences are lighter every couple of weeks, and they are minimal. He will start from now on trying to put a bit more data on HEP-SCORE. Helge commented that it is nice to see that this steady progress on in quite a number of sides.

Domenico reported that the Juno software experts confirmed that 2.7 GB per core memory consumption is needed in the initial phase out the workload. About the missing workloads that will enter in the script D, the gravitational waves one is ready. The container is working fine and the benchmarking working group members will receive the notification to start the validation, before making it available for all the sides. The ALICE workload is in a quite advanced status. If ALICE workload is ready, we will include in the script D both ALICE and GW workloads. Otherwise, we will have a script D with GW workload and the script E will be created when the ALICE one is fine.

Stefano explained that ALICE workload is working fine in grid being in production, but there was a failure in digitization step when the workload runs in the HEP-SCORE container. ALICE experts succeeded in debugging the interplay of the workload with the container environment.

Helge commented that it's good that this issue has been isolated and it's also appreciated that this is being followed up with high priority, but Helge remarked his worries on timescales because the week after Easter possibly a number of people will not be available. It would also be a pity to conclude that we need to move ahead without ALICE or if waiting for these workloads would delay in a very significant way.

Any other business

Helge reminded that in two weeks' time we have not scheduled a meeting. The week after there is HEPiX workshop, where Michele, Domenico and Helge have decided to throw in 2 abstracts where the boundary is more strategy towards HEP-SCORE and the technical issues. Michele will cover the technical issues including the results that Randy has presented during the present meeting. Helge will present the strategy where we are, the proposal for a new benchmark and the accompanying transition scenario from HS06 to the new benchmark.

Next meeting

Wed 04 May at 14:00 h UTC (16:00 h in Geneva)

Annex: Attendance

Present:

Miltiadis Alexis (CERN)
Tommaso Boccali (INFN Pisa)
Domenico Giordano (CERN)
Michel Jouvin (IJCLab)
Walter Lampl (U Arizona)
Helge Meinhard (CERN, chair)
Andrew Melo (Vanderbilt U)
Gonzalo Menendez Borge (CERN; notes)
Gonzalo Merino (PIC)
Stefano Piano (INFN Trieste)
Andrea Scialba (CERN)
Randall Sobie (U Victoria)
Andrea Valassi (CERN)
Tony Wong (BNL)
Yan Xiaofei (IHEP)