

Summary of Benchmark Workshop

Randall Sobie University of Victoria
(on behalf of the HEPiX Benchmark Working Group)

HEPScore

Recall the motivation for HEPSScore

HEPSpec06 is old and no longer supported

Some experiment workloads give similar results to HEPSpec06 but not all experiments

HEPScore shows that the workloads are more performant on new CPU architectures

HEPScore (CPU) will help us develop HEPSScore (CPU+GPU)

Benchmark Suite

Should future benchmark runs at sites be done via the suite?

Positive feedback from the site reports

Already helping with procurement

Envisage the WLCG sites running the Suite with the data (and its metadata) written to Elastic Search at CERN

Sites make their own calculation of HEPSpec06-hours (expect the same for HEPScore)

Tuesday accounting session:

- Sites could run the Suite for new software or after major upgrades (e.g. OS upgrades)
- Sites with "older CPUs" could use a look up table

Workloads

Workload	Running Time (m)	# of events * # of threads
Atlas_gen_sherpa	31	200 * 1
Atlas_reco_mt	69	100 * 4
Atlas_sim_mt	156	5 * 4
CMS_gen_sim	42	20 * 4
CMS_digi	31	50 * 4
CMS_reco	51	50 * 4
Belle2_gen_sim_reco	25	50 * 1
Alice_gen_sim_reco	194*	3 * 4
LHCb_gen_sim	104	5 * 1
Juno_gen_sim_reco	67	50 * 1
Gravitational Wave	138	1 * 4
Total	908 (15+ hours)	

Times for three runs on reference machine

* - Alice reco currently not included in benchmark score, due to technical problems with reco workload. Reco is ~ 50% of running time. Once issue is resolved, could run only reco to shorten workload length.

Are Experiments ready to freeze their workloads today?

Most experiments indicated their workloads are ready

Need to finalize workloads in the next months
(more on ALICE on later slide)

Need to review how to manage changes to workloads in future years - Run4 may bring significant changes

HEPScore

□ Workloads and weights to be used in HEPscore

Workload	Running Time (m)	# of events * # of threads
Atlas_gen_sherpa	31	200 * 1
Atlas_reco_mt	69	100 * 4
Atlas_sim_mt	156	5 * 4
CMS_gen_sim	42	20 * 4
CMS_digi	31	50 * 4
CMS_reco	51	50 * 4
Belle2_gen_sim_reco	25	50 * 1
Alice_gen_sim_reco	194*	3 * 4
LHCb_gen_sim	104	5 * 1
Juno_gen_sim_reco	67	50 * 1
Gravitational Wave	138	1 * 4
Total	908 (15+ hours)	

Times for three runs on reference machine

* - Alice reco currently not included in benchmark score, due to technical problems with reco workload. Reco is ~ 50% of running time. Once issue is resolved, could run only reco to shorten workload length.

Possible HEPscore-7 candidate:

ATLAS gen_sherpa and reco	100m
CMS gen_sim and reco	93m
Belle2	25m
LHCb	104m
ALICE reco	50m (?)

Estimated time **400m (7 hours)**

Not in HEPscore-7:

Juno, Gravity Wave, CMS_digi and ATLAS_sim_mt

Weighting

□ Workloads and weights to be used in HEPscore

Workload	Running Time (m)	# of events * # of threads
Atlas_gen_sherpa	31	200 * 1
Atlas_reco_mt	69	100 * 4
Atlas_sim_mt	156	5 * 4
CMS_gen_sim	42	20 * 4
CMS_digi	31	50 * 4
CMS_reco	51	50 * 4
Belle2_gen_sim_reco	25	50 * 1
Alice_gen_sim_reco	194*	3 * 4
LHCb_gen_sim	104	5 * 1
Juno_gen_sim_reco	67	50 * 1
Gravitational Wave	138	1 * 4
Total	908 (15+ hours)	

Times for three runs on reference machine

* - Alice reco currently not included in benchmark score, due to technical problems with reco workload. Reco is ~ 50% of running time. Once issue is resolved, could run only reco to shorten workload length.

“Equal weighting” is similar to the “WLCG weighting” for HEPscore-7

ATLAS (28%), CMS (28%)
ALICE, Belle2, LHCb 15% each

Overweighting B2

Keep otherwise HEPscore is only LHC experiments

But it was shown that changing the weightings had minimal impact

ALICE reco

Want to include ALICE reco (Pb-Pb events)

ALICE-gen-sim similar results to other LHC experiments

Proposal

HEPiX WG could take the current ALICE workload (gen-sim-reco) and pin the CPU usage

Quickly validate the reco workload

Allow time to develop a standalone ALICE-reco workload

Outlook

Converge on HEPScore (workloads and weighting)

- HEPScore-7 and equal weighting (Grid weighting)
- Finalize alice_reco workload and cross check results

Identify a number of production sites to run the Suite

Test configuration:

HEPSpec06 + HEPScore-7 + Other workloads (not in HEPScore-7) 1 day to run

Find out how sites report their results (single HEPSpec06 or weighted average of CPU models)

Write results to Elastic Search

Ongoing R&D effort to determine a CPU+GPU benchmark and studies of power usage