HEPscore benchmark status

D. Giordano (CERN/IT)

on behalf of
HEPiX CPU Benchmarking WG & HEPscore Deployment TF
hepix-cpu-benchmark@hepix.org

WLCG Accounting TF
25 Nov 2021

HEP Benchmarks project goal

- Identify a replacement of HS06 for CPUs

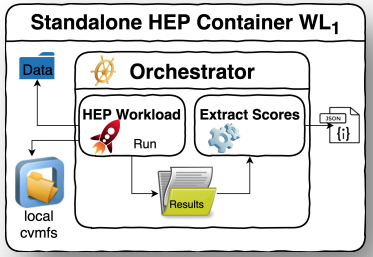
- Reasons
  - HS06 is not supported anymore by SPEC (since 2018)
  - Signs of discrepancies w.r.t. the “run2” LHC Experiments’ applications
    - Could be worse with run3 applications
  - HS06 is not a HEP-specific benchmark
    - Desire to apply the same benchmark to other architectures (ARM, GPUs, …)

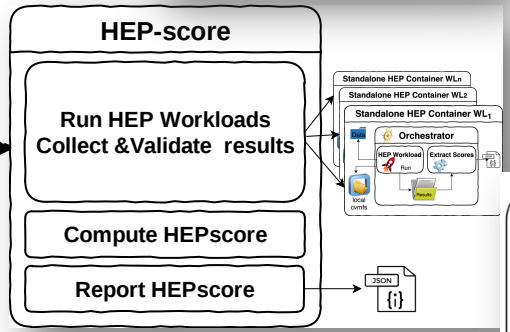
- HEPscore is proposed by the HEPiX Benchmarking WG
  - Uses the workloads of the experiments as application benchmarks
  - Combine them in a single score as HS06 does

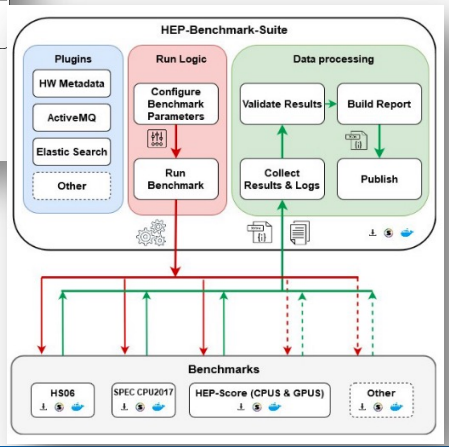
HEP Benchmarks project

Three main components

- *HEP Workloads*
  - Individual **reference** HEP applications
  - Common build infrastructure
- *HEP Score*
  - Orchestrate the run of a series of HEP workloads
  - Compute the **HEPscore** value
  - Report whole set of WL results
- *HEP Benchmark Suite*
  - Orchestrator of multiple benchmark suites
    - HEPscore, HS06, SPEC CPU2017…





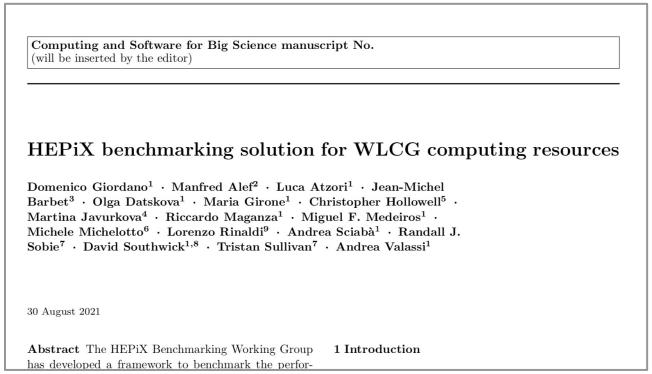


Achievements

- HEP Workloads “containeraized” and fully validated
  - LHC Experiments (Run2 version) + Belle II
    - All production steps (Gen, Sim, Digi, Reco) available
  - GPU workloads SimpleTrack (LHC simulation)
- HEPscore v1.2 released
  - Singularity & Docker supported, improved report, Python wheels available
  - Default config validated up to 256 cores
- HEP Benchmark Suite v2.1 released
  - Metadata section with detailed HW information, install as unprivileged user, python wheels available
- HS06 extended to ARM CPUs
- HEPiX Spring presentation & Demo

Achievements (II)

- Analysis of HEPscore Vs HS06 measurements
  - Using the ***demonstrator*** benchmark, $HEPscore_{\beta}$,
  - Outcome: *it may be possible to create a new benchmark for CPUs based on HEP applications*
    - Presented in the Autumn HEPiX workshop
    - Publication accepted by *Computing and Software for Big Science*
      - "*HEPiX benchmarking solution for WLCG computing resources*"
- Positive feedback from early adopters



Computing and Software for Big Science manuscript No.
(will be inserted by the editor)

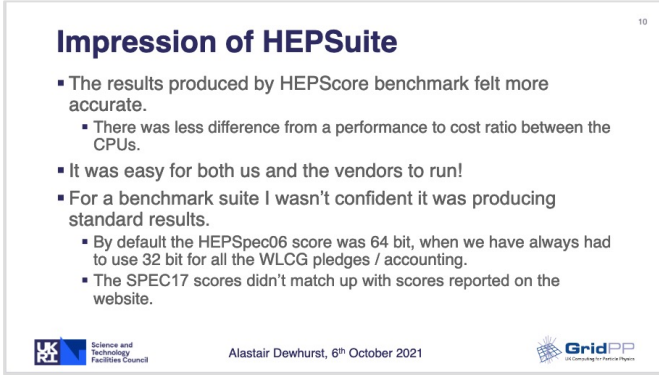
HEPiX benchmarking solution for WLCG computing resources

Domenico Giordano[1] · Manfred Alef[2] · Luca Atzori[1] · Jean-Michel Barbet[3] · Olga Datskova[1] · Maria Girone[1] · Christopher Hollowell[5] · Martina Javurkova[4] · Riccardo Maganza[1] · Miguel F. Medeiros[1] · Michele Michelotto[6] · Lorenzo Rinaldi[9] · Andrea Sciabà[1] · Randall J. Sobie[7] · David Southwick[1,8] · Tristan Sullivan[7] · Andrea Valassi[1]

30 August 2021

Abstract The HEPiX Benchmarking Working Group has developed a framework to benchmark the perfor-

1 Introduction



Current work

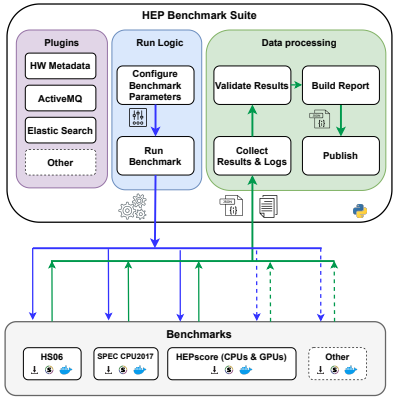
- Inclusion of new WLCG workloads in the HEP Workloads catalogue
  - In collaboration with the HEPscore deployment Task Force
  - Identified WLs other than LHC experiments
- Start a campaign of measurements
  - Multiple CPU models from several sites
  - Multiple benchmarks (HS06, SPEC CPU 2017, HEPscore)
- Objective
  - Compare HEP WL performance, identify representative subset for what will be HEPscore2X configuration
  - Compare performance w.r.t. SPEC benchmarks

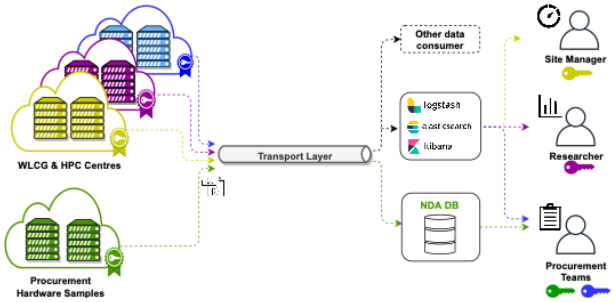
WL	Responsible	OS	Platform	WL developed in a git fork (if relevant)	Merged in HEP-Workloads repo	Built	Validated	Reference score
Alice Sim	S. Piano	cc7	x86					
Atlas simMT	W. Lamp	cc7	x86					
LHCb gen-sim 2021	A. Valassi	cc7	x86					
CMS gen-sim Run3	A. Sciabà	cc7	x86/arm			x86		
CMS Digi Run3	A. Sciabà	cc7	x86/arm			x86		
CMS Reco Run3	A. Sciabà	cc7	x86/arm			x86		
Belle2	R. Sobie	cc7	x86					
Dune	A. Mc Nab	cc7	x86	https://gitlab.cern.ch				
Juno	X. Yan	cc7	x86	https://gitlab.cern.ch				
Grav-Wave	J. Willis	cc7	x86	https://git.ligo.org/j				
Madgraph	A. Valassi	cc7	x86 / GPU					

Legend	Latest changes are in lighter colours
Ok	
In progress	
Blocker	
Not started yet	

HEP Benchmark Suite

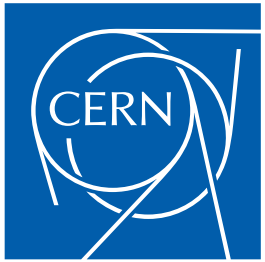
- Is the toolkit to run and collect results centrally
  - Execution of multiple benchamrks; document validation; metadata enrichment
- Used @ CERN in the enrollment of all servers. Several benchmark scores are collected an monitored





Conclusions

- The HEP Benchmarks project is progressing well
- The composition of HEPscore2x will be likely defined in 2022
  - After approval of WLCG, it's reasonable to envisage a period of coexistence of HS06 and HEPscore benchmarks
- Feedback and support request welcome in the dedicated Discourse Forum and gitlab issues tracker


CERN

Benchmark comparing “speed factors”

$$\bar{x} = \left(\prod_{i=1}^{n} x_i^{w_i} \right)^{1/\sum_{i=1}^{n} w_i}$$

https://en.wikipedia.org/wiki/Weighted_geometric_mean

- In order to compare servers **HS06** and **HEP-Score** implement the **geometric mean** approach. Needs:
  - a set of reference workloads (**WLs**)
  - a measure of performance per WL (m_i), that typically goes as [1/s] (eg. can be the event throughput)
  - a reference machine
- The score **S** of a server (**srv**) is defined as the **geometric mean** of the **speed factors** x_i(**srv**,ref) = m_i(**srv**)/m_i(**ref**) respect to the reference machine (**ref**)
  - i.e. “speed” is *normalised* respect to the reference machine “speed”
- The relative score between srv_A and srv_B is the ratio of the scores S(srv,ref) , this is still a geometric mean of speed factors

	WL_1		WL_2		WL_n		Score	S(A,B)
Ref. Srv 	m_1(ref)	1 (by def)	m_2(ref)	1 (by def)	m_n(ref)	1 (by def)	$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$	
Srv A 	m_1(A)	x_1(A,ref)	m_2(A)	x_2(A,ref)	m_n(A)	x_n(A,ref)	S(A,ref)	$\frac{S(A,ref)}{S(B,ref)}$
Srv B 	m_1(B)	x_1(B,ref)	m_2(B)	x_2(B,ref)	m_n(B)	x_n(B,ref)	S(B,ref)	