CPU Benchmarks: update

D. Giordano (CERN/IT)
on behalf of
HEPiX Benchmarking WG
WLCG HEPscore deployment TF

GDB
9 March 2022

Starting point

- Identify a replacement of HS06 as CPU benchmark

- Reasons
  - Since 2018 HS06 is **not supported** anymore by the SPEC corporation
  - Signs of **discrepancies** w.r.t. the Run 2 LHC experiments' applications
    - Need to investigate scaling w.r.t. new (Run 3) HEP applications
  - HS06 is not a HEP-specific benchmark
    - Interest in adopting a **field-specific benchmark**
  - Desire to identify a benchmark suitable for all **architectures** to be adopted in HEP (x86, 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

Two teams working in strict contact

HEPiX Benchmarking WG

- Roles
  - Evaluation of benchmark alternatives
  - Design and development of the **HEP Benchmarks project**
  - Validation of the HEP workloads
  - Analysis of benchmark measurements
- The team(*), ~13 people, meets weekly
- Active (again) since 2018
- Chairs: Michele Michelotto (INFN) & D.G.

WLCG HEPscore deployment TF

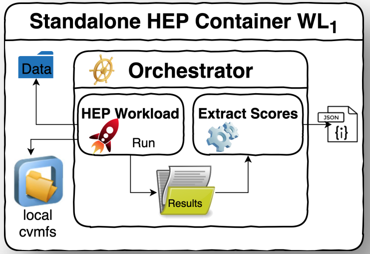
- Roles
  - Propose a migration scenario from HS06 to the new benchmark
    - Recommend the benchmark composition
    - **Primary focus on x86 arch**
  - Coordinate the collection of recent workloads from Experiments
  - Onboard WLCG sites in the benchmark measurements
- The team(*), ~20 people, meets bi-weekly
- Started on Nov 4. 2020
- Chair: H. Meinhard (CERN)

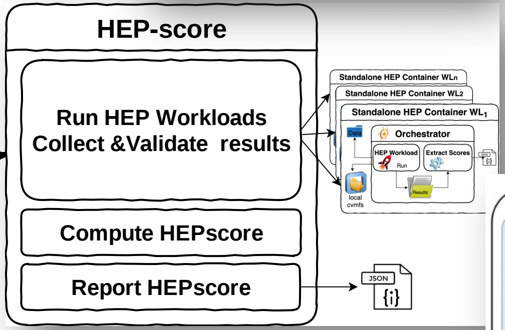
(*) See backup slides for member list

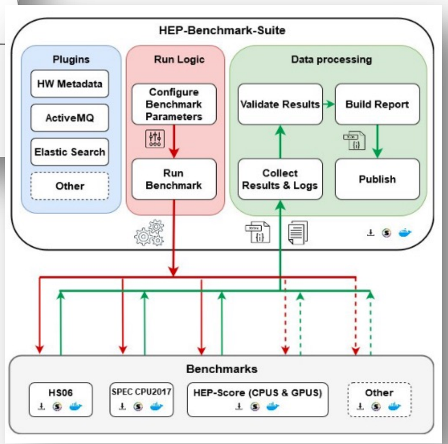
HEP Benchmarks project

Three main components

- *HEP Workloads*
  - **Standalone reference** HEP applications
    - No network needed at runtime to access sw and data
  - Common build infrastructure
- *HEP Score* (the benchmark)
  - Orchestrate the run of a series of HEP workloads
  - Compute the **HEPscore** value
  - Report whole set of WL results
- *HEP Benchmark Suite*
  - Framework to run likewise multiple benchmarks
    - HEPscore, HS06, SPEC CPU2017…







Recap WG milestones: Spring 2021

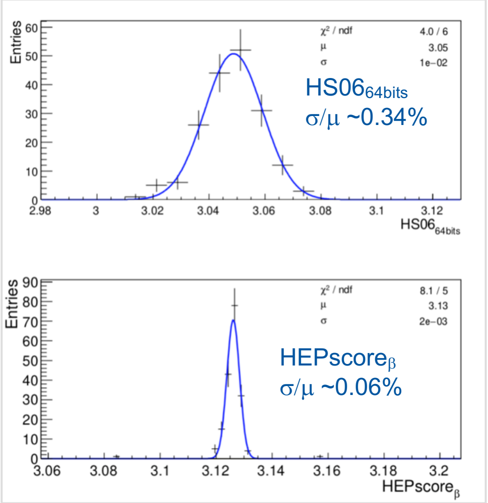
Main subjects in the HEPiX Spring presentation & Demo report of the WG

- ❑ LHC Run 2 Workloads (+ Belle2) “containeraized” and fully validated
  - All production steps (Gen, Sim, Digi, Reco) available
  - Limited GPU workloads: only SimpleTrack (LHC simulation)
- ❑ HEPscore v1.2 released
  - Singularity & Docker supported; Python wheels available
  - Default config ($HEPscore_{\beta}$) validated up to 256 cores
- ❑ HEP Benchmark Suite v2.1 released
  - Metadata section with detailed HW information, install as unprivileged user, python wheels available

Recap WG milestones: Fall 2021

Main subject in the Autumn HEPiX report of the WG

- ❑ Analysis of HEPscore$_\beta$ Vs HS06 measurements
  - – Published in the CSBS paper (Springer journal)
    HEPiX benchmarking solution for WLCG computing resources
    - Outcome: "*it may be possible to create a new benchmark for CPUs based on HEP applications*"
    - Obtained with the ***demonstrator*** benchmark, HEPscore$_\beta$, based on Run 2 HEP workloads



- ❑ Positive feedback from the beta testers of HEPscore

TF activity outcomes in 2021

Prior to any recommendation about the new benchmark composition, the TF agreed on

- ❑ **Extending** the list of **standalone** HEP workloads
  - – Run3 workloads from LHC experiments
  - – Onboard other HEP communities that currently are using HS06
    - • Belle2, Dune, Juno, Grav. Waves Exp (IGWN)
- ❑ **Studying** the workloads' performance on several servers
  - – Cover as much as possible all the CPU models deployed in production in WLCG
  - – Include SPEC CPU benchmarks
    - • HS06 as well as SPEC CPU 2017 (**intrate** and **cpp** configs)

"matrix" of score measurements

	CPU model A	**CPU model B**	**…**
$HS06_{64bit}$			
$HS06_{32bit}$			
SPEC CPU 2017 (cpp config)			
SPEC CPU 2017 (intrate config)			
HEP WL_1			
HEP WL_2			
…			
HEP WL_n			

Action 1: Extend list of HEP workloads

- The TF to coordinate the new workloads' identification
- The WG to build the **standalone** containers
  - In strict contact with the experiments' experts
- 10 workloads for **x86** to enter in the **matrix**
  - 8 ready; 4 are Single Process or Single Thread
- In addition, 2 **prototype** workloads for GPU (Madgraph generator and CMS HLT-like)
  - Demonstrate the HEPscore usability on other arch. (**longer term objective!**)

Legend
Ok
In progress
Not started yet

	A	B	C	D	E	F	G	H	I	J
1	WL	Responsible	OS	Platform	WL developed in a git fork (if relevant)	Merged in HEP-Worklaods repo	Built	Validated	Reference score	Ready for the "metrix"
2	Alice Gen-Sim-Reco	S. Piano	cc7	x86						
3	Atlas gen sherpa	W. Lamp	cc7	x86						
4	Atlas simMT	W. Lamp	cc7	x86						
5	LHCb gen-sim 2021	A. Valassi	cc7	x86						
6	CMS gen-sim Run3	A. Sciabà	cc7	x86/arm			x86/arm			
7	CMS Digi Run3	A. Sciabà	cc7	x86/arm			x86/arm			
8	CMS Reco Run3	A. Sciabà	cc7	x86/arm			x86/arm			
9	CMS HLT-like	A. Sciabà	cc7	x86 & GPU			x86			
10	Belle2	R. Sobie	cc7	x86						
11	Dune	A. Mc Nab	cc7	x86	https://gitlab.cern.			On hold for lack of time from Dune experts		
12	Juno	X. Yan	cc7	x86	b.cern.ch/xiaofei/he					
13	Grav-Wave	J. Willis	cc7	x86	https://git.ligo.org/					
14	Madgraph	A. Valassi	cc7	x86 / GPU						

Highlights of the new HEP workloads

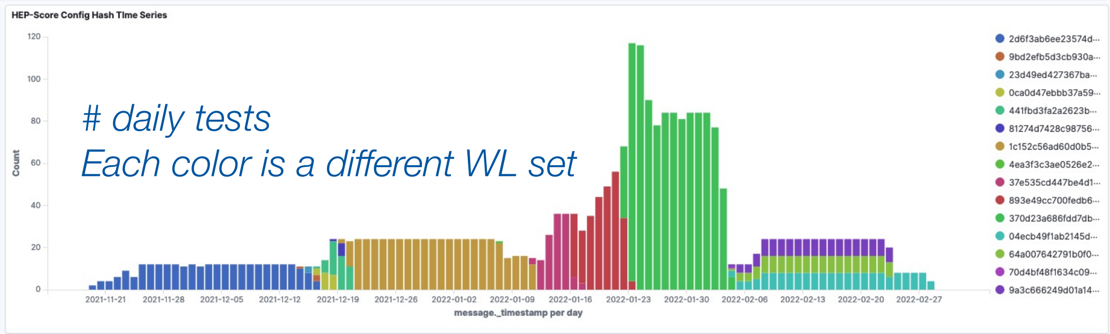
- Alice: gen-sim-reco based on Run3 Online-Offline (**O2**) framework
- Atlas: new gen WL uses **Sherpa** as generator. Sim WL is **multi-threaded**
- CMS: the standalone WLs are distributed also for **ARM** platforms (a nice to have). All WLs are multi-threaded
- Juno: required improvement of the HEP-Workloads infrustructure to snapshot a different cvmfs endpoint
- IGWN is not an event based workload
- *All Singularity SIF images are distributed via the Harbor registry at CERN*
  - Speedup the pre-run phase: faster download, conversion from Docker images not needed anymore

Legend
Ok
In progress
Not started yet

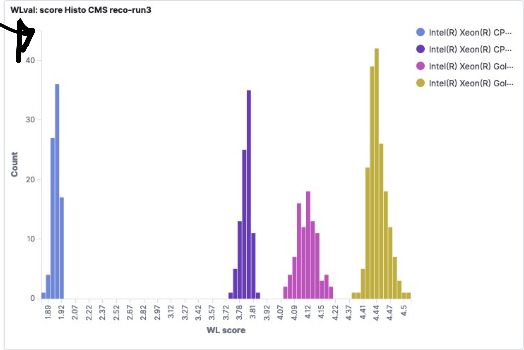
	A	B	C	D	E	F	G	H	I	J
1	WL	Responsible	OS	Platform	WL developed in a git fork (if relevant)	Merged in HEP-Worklaods repo	Built	Validated	Reference score	Ready for the "metrix"
2	Alice Gen-Sim-Reco	S. Piano	cc7	x86						
3	Atlas gen sherpa	W. Lamp	cc7	x86						
4	Atlas simMT	W. Lamp	cc7	x86						
5	LHCb gen-sim 2021	A. Valassi	cc7	x86						
6	CMS gen-sim Run3	A. Sciabà	cc7	x86/arm			x86/arm			
7	CMS Digi Run3	A. Sciabà	cc7	x86/arm			x86/arm			
8	CMS Reco Run3	A. Sciabà	cc7	x86/arm			x86/arm			
9	CMS HLT-like	A. Sciabà	cc7	x86 & GPU			x86			
10	Belle2	R. Sobie	cc7	x86						
11	Dune	A. Mc Nab	cc7	x86	https://gitlab.cern.			On hold for lack of time from Dune experts		
12	Juno	X. Yan	cc7	x86	b.cern.ch/xiaofei/he					
13	Grav-Wave	J. Willis	cc7	x86	https://git.ligo.org/					
14	Madgraph	A. Valassi	cc7	x86 / GPU						

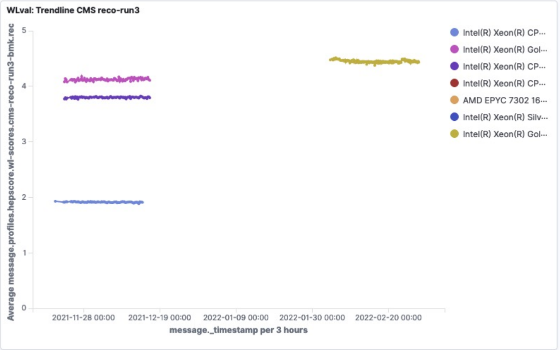
Validation of the new HEP Workloads

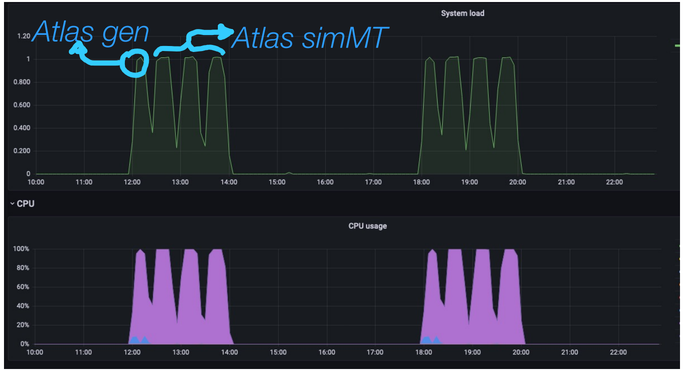
- Before release, continuous runs on CERN testbed of 3 servers
  - Validate the WL robustness (no failures) and stability (i.e. benchmark resolution)
- Multiple monitoring dashboards
  - Follow progress, inspect results
  - Verify that the servers are fully loaded



CMS reco score

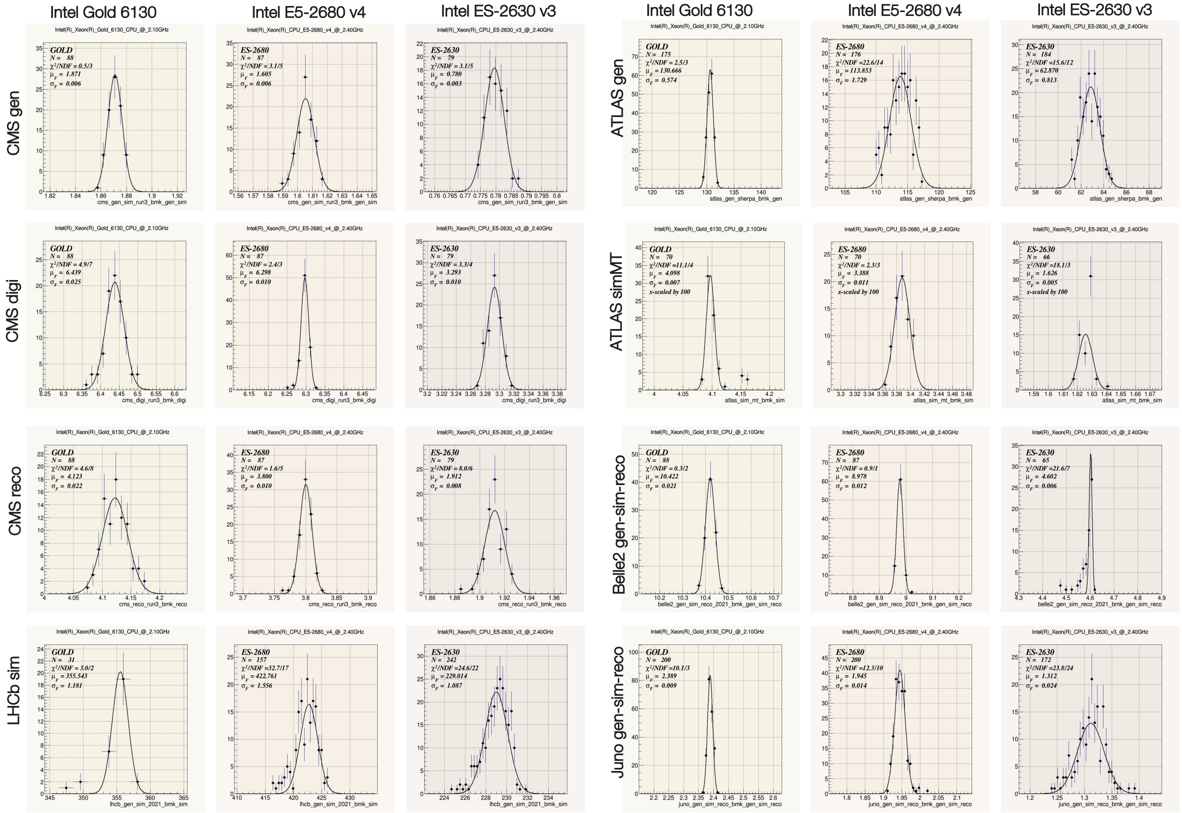






Validation of the new HEP Workloads (II)

- Offline analysis to complete and confirm the online monitoring
  - CERN testbed of 3 servers
- Quantified the WL score stability
  - Standard deviations of measurements typically < 1%



Action 2: Run on multiple x86 platforms

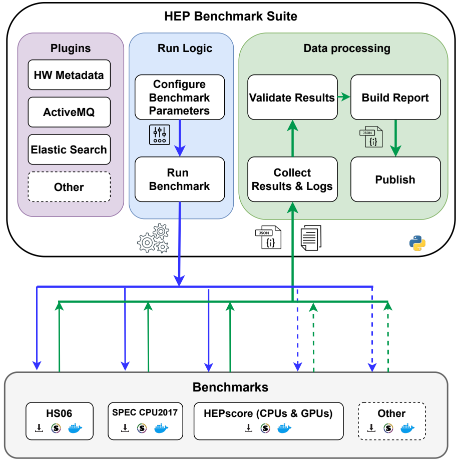
- ❑ TF to promote the identification of a set servers representative of the WLCG production platforms
  - – Encourage WLCG sites to offer servers and expertise to run the benchmarks
  - – And supervise the progress of the benchmark process

- ❑ WG to assist volunteers in the benchmark execution
  - – Scripts to run the HEP Benchmark Suite
  - – Dashboards to monitor the data collection
  - – Gather feedback for improvements

"matrix" of score measurements

	CPU model A	**CPU model B**	…
$HS06_{64bit}$			
$HS06_{32bit}$			
SPEC CPU 2017 (cpp config)			
SPEC CPU 2017 (intrate config)			
HEP WL_1			
HEP WL_2			
…			
HEP WL_n			

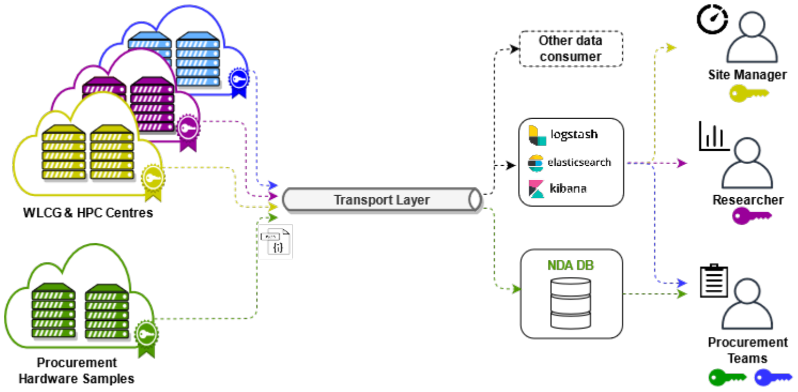
How WLCG sites run benchmarks



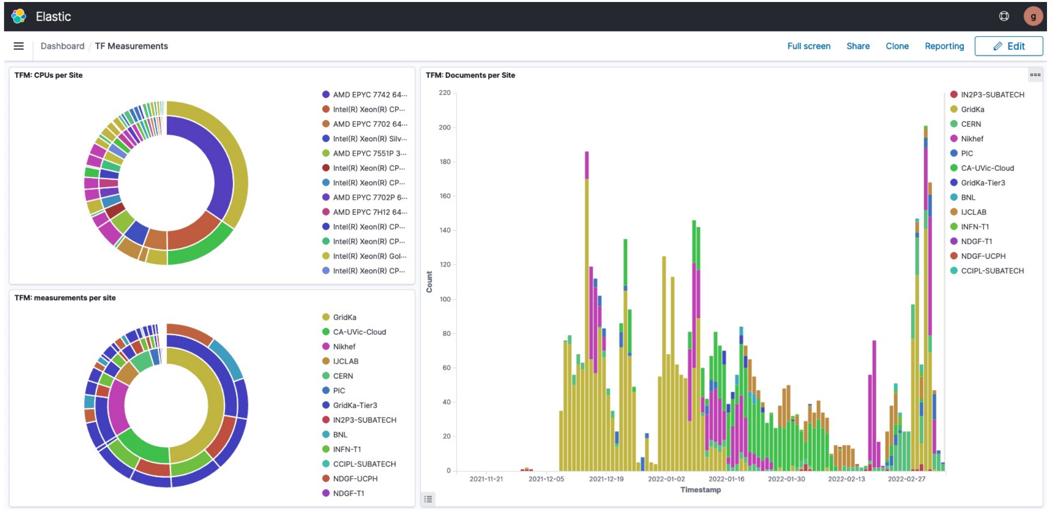
- A **single script** for each benchmark configuration, to trigger the Suite execution:
  - Run benchmark -> Extract server metadata -> Validate the overall report -> Publish on the remote Elasticsearch DB
- The progress so far

succeeded
ongoing
problem
No SPEC licence

Site	HS06_32	HS06_64	SPEC17_Intrate	SPEC17_cpp	HEP WLs A	HEP WLs B	HEP WLs C	HEP WLs D
BNL								
CA-UVic-Cloud								
IHEP								
IJCLab								
KIT (Gridka)								
LIGO								
NDGF-T1								
Nikhef								
PIC								
RAL								
INFN-T1								
SUBATECH								



SPEC benchmark measurements

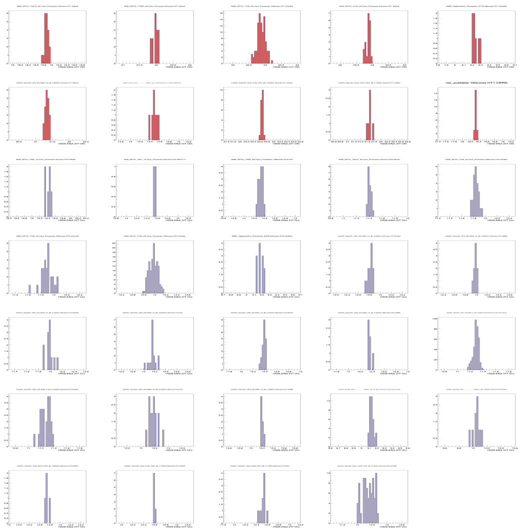


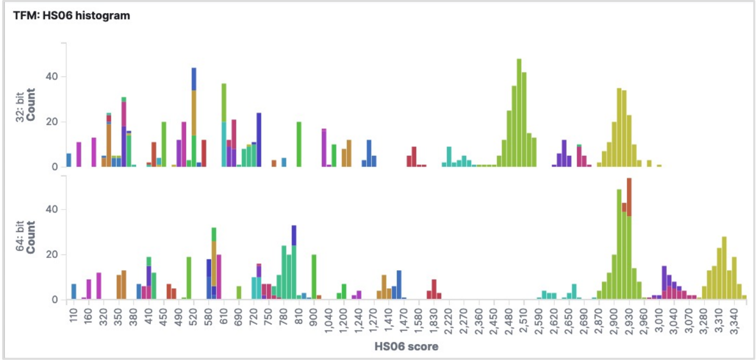
Processor	# Sites	# SMT configs	# RAM config
Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz	3	1	2
AMD EPYC 7302 16-Core Processor	2	1	2
AMD EPYC 7702 64-Core Processor	2	2	2
Intel(R) Xeon(R) CPU E5520 @ 2.27GHz	2	2	2
Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz	2	1	2
Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz	2	1	2
Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz	2	1	2
Intel(R) Xeon(R) Gold 5118 CPU @ 2.30GHz	2	3	1
Intel(R) Xeon(R) Silver 4210 CPU @ 2.20GHz	2	1	1
Intel(R) Xeon(R) Silver 4216 CPU @ 2.10GHz	2	2	2
AMD EPYC 7351 16-Core Processor	1	1	1
AMD EPYC 7443P 24-Core Processor	1	1	1
AMD EPYC 7452 32-Core Processor	1	1	1
AMD EPYC 7551P 32-Core Processor	1	2	1
AMD EPYC 7702P 64-Core Processor	1	1	1
AMD EPYC 7713 64-Core Processor	1	1	1
AMD EPYC 7742 64-Core Processor	1	2	1
AMD EPYC 7H12 64-Core Processor	1	1	1
AMD Opteron(tm) Processor 6174	1	1	1
AMD Opteron(tm) Processor 6376	1	1	1
Intel Core Processor (Haswell, no TSX, IBRS)	1	1	1
Intel(R) Xeon(R) CPU E5630 @ 2.53GHz	1	1	1
Intel(R) Xeon(R) CPU E5-2640 v3 @ 2.60GHz	1	1	1
Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz	1	1	1
Intel(R) Xeon(R) CPU E5-2665 0 @ 2.40GHz	1	2	2
Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz	1	3	2
Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz	1	1	1
Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz	1	1	1
Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz	1	1	1
Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz	1	1	1
Intel(R) Xeon(R) Gold 6252 CPU @ 2.10GHz	1	1	1

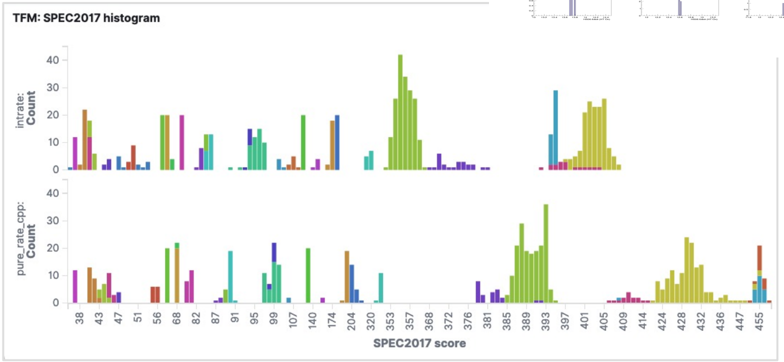
- ~5k measurements collected since Dec ’21
- >30 distinct CPU models from 11 WLCG sites
  - Multiple HT and RAM configurations available
  - Same model in multiple sites: only in 1/3 of the cases

SPEC-family benchmarks

- Executed benchmarks a significant number of times (>10) to perform accurate studies
- Detailed data analysis
  - by extracting data from ES
  - offline data cleaning and fits





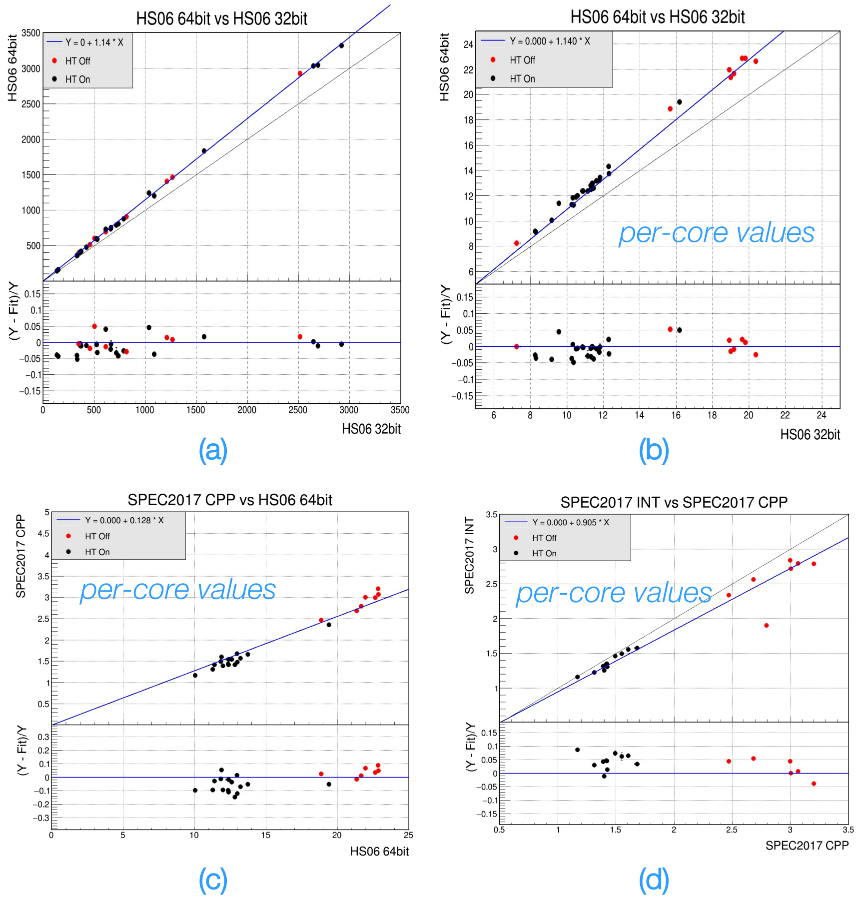


First (preliminary) study

SPEC values across the benchmarked CPU models

- (a,b) **Confirm** the average conversion factor between $HS06_{32}$ and $HS06_{64}$ is **1.14** (i.e. $HS06_{64}$ is 14% higher than $HS06_{32}$)
- (c) **Confirms** the **correlation** of $HS06_{64}$ and SPEC CPU2017 **cpp rate** config., as already reported by the **WG** in the past (see paper)
- (d) Shows that SPEC CPU 2017 **intrate** and **cpp** config. are **compatible** within 10%

NB: (b,c,d) data are normalized to the number of active cores



Sites can still contribute to the measurements

Request to MB Members (1)

- Contact sites that you work with:
  - Compare our list with their work force for batch processing; any major contributions by configurations not in the list? (*)
  - Any configurations by that site already proposed by a different site? (*)
  - Any configurations by that site on the list? Please confirm that bulk processing uses the same configuration (**)

22-Feb-2022 Helge Meinhard (at) CERN.ch – HEP-SCORE deployment task force

Request to MB Members (2)

(*) Please consider proposing one such configuration for benchmarking:

- Material: One server, not necessarily available for benchmarking all the time, but should be exclusively used for benchmarking during campaigns
- Personnel: One contact person needed for basic babysitting – not very demanding nor time-consuming

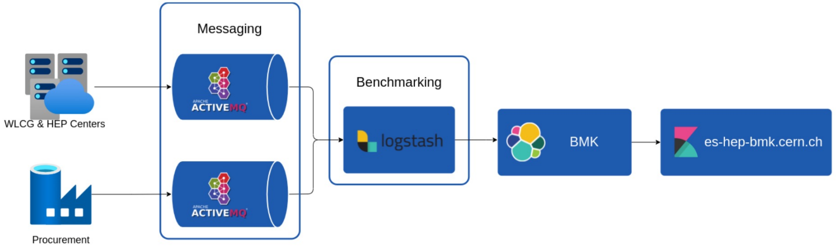
(**) Configurations include type and number of processors, HT/SMT settings, number of SMT cores used, RAM per core; workload running on bare metal, in containers, in VMs (what size)?

22-Feb-2022 Helge Meinhard (at) CERN.ch – HEP-SCORE deployment task force 9

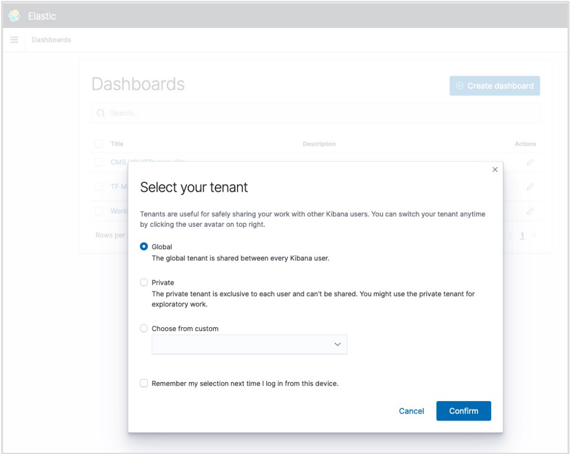
- ❑ Detailed server configuration in the Helge's presentation to the MB
- ❑ Contact *hep-benchmarks-support@cern.ch*

Consolidation of the ES infrastructure (Fall 2021)

- Elasticsearch is the storage solution selected for the monitoring, analysis and long-term preservation of the benchmark results



- In September 21 decided to adopt the new CERN ES infrastructure based on Open Distro
  - Take advantage of the new features offered by Open Distro
  - Implied migration from previous ES: data collected, and dashboards developed in the past years



- Separated tenants coexist in the same cluster to access procurement data or WLCG data
  - ACLs in place for authorization access. Based on egroups.
    - Still few issues to solve: individual cases of externals not having access via their CERN/Edugain certificate & SSO

Next steps

- Workloads:
  - Finalize development and validation of Alice and IGWN workloads (end of March)
- Measurements:
  - Complete the “matrix” (end of April)
- Analysis:
  - Study the relative scaling of the HEP workloads w.r.t. the available CPU models, as started for the SPEC family (April-May)
  - Identification of the HEP workloads’ set that will define HEPscore22 (optimistically in Q3)
    - Possibly avoid to run correlated WLs, using **weights** in the HEPscore22 definition

Conclusions

- ❑ Efficient collaboration between two teams
  - – WLCG HEPscore deployment TF
  - – HEPiX Benchmarking WG
- ❑ Large coverage of recent HEP software applications
  - – Run 3 LHC experiments + other HEP experiments
  - – The standalone containers are robust and provide reproducible results
- ❑ Numerous WLCG sites contribute substantially with test systems on which benchmarks are run
  - – More contributions are needed and encouraged!!
- ❑ A great "Thank You!" goes to the experts from the experiments and the sites as well as to the WG and TF members

(*) See backup slides for member list

CERN

WLCG TF members

Experiments' experts and/or site representatives and/or WLCG MB members:

Manfred Alef (KIT), Miltiadis Alexis (CERN), Tommaso Boccali (INFN Pisa), Simone Campana (CERN), Ian Collier (STFC-RAL), Alastair Dewhurst (STFC-RAL), Domenico Giordano (CERN), Michel Jouvin (IJCLab), Walter Lampl (U Arizona), Helge Meinhard (CERN, chair), Andrew Melo (Vanderbilt U), Gonzalo Menendez Borge (CERN), Gonzalo Merino (PIC), Bernd Panzer-Steindel (CERN), Randall Sobie (U Victoria), Stefano Piano (INFN Trieste), Matthias Schnepf (KIT), Oxana Smirnova (U Lund), Jeff Templon (Nikhef), Andrea Valassi (CERN), Josh Willis (Caltech), Tony Wong (BNL), Yan Xiaofei (IHEP)

HEPiX Benchmarking WG members

Manfred Alef (KIT), Luca Atzori (CERN), Jean-Michel Barbet (IN2P3-Subatech), Matthew Franklin Ens (UViC), Domenico Giordano (CERN), Christopher Henry Hollowell (BNL), Matthias Schnepf (KIT), Gonzalo Menendez Borge (CERN), Michele Michelotto (INFN), Andrea Sciaba' (CERN), Randy Sobie (UViC), David Southwick (CERN), Andrea Valassi (CERN)