WLCG HEP-SCORE Deployment Task Force

Helge Meinhard / CERN-IT
Online HEPiX
16 March 2021

Why a CPU Benchmark?

- From WLCG perspective, most importantly
    - Experiment requests and site pledges
    - Accounting of CPU usage

- Many sites also use them for procurements

CPU Benchmarks in WLCG

- Pre-2009: WLCG used a benchmark based on SPECint 2000 (part of SPEC-CPU 2000)

- At EOL, WLCG started looking into SPEC CPU 2006
  - C++ applications of both SPECint 2006 and SPECfp 2006 matched applications well

- Defined benchmark in January 2009 as **HEP-SPEC06**

- At EOL, WLCG started looking into SPEC CPU 2017
  - Found to be a sub-optimal match

HEP-SPEC06

- Defining workload is not enough
  - Also need to describe the conditions of running
  - Chose conditions in 2009 that were as realistic as possible in view of CPU farms in use for WLCG
    - gcc version and flags, 32bit app, as many concurrent processes as cores, …
  - Scaling behaviour of real workload initially well within about 10% of benchmark

  - Conditions have changed much since then
    - 64bit, new compilers/versions, flags, SMP processors, multi-threaded applications or pilots launching identical binaries multiple times, VMs and containers, …
  - Still maintained the initial choices
    - Scaling behaviour matching real workload still surprisingly well (often ~ 20% or better with some exceptions)

HEP-SPEC06: A Success Story

- Used successfully for more than ten years

- Initially designed as a tool for WLCG, found widespread use in other communities (not limited to HEP)

- Key reasons IMO:
  - Benchmark defined as one single number
  - Definition did not change during the HEP-SPEC06 lifetime

HEP-SPEC06 Criticism

- Individual reports of scaling deviations of 40% and more
  - Some suggested replacements turned out to be worse for typical workload mixes
- Benchmark workload not typical of HEP applications
  - Would the reasonable scaling persist with non-x86 CPUs, for example? Well, we didn't have that issue (yet)...
- Running HEP-SPEC06 requires a software licence from SPEC
  - Strong desire to consider licence-free benchmarks as successor

- Not representative of full machine potential, not representative of improved experiment workload
  - These are "features" rather than "bugs"

- Time to move on…

Future HEP Benchmarking (1)

- Pretty much like in 2007/2008, benchmark experts got together and worked on a new benchmark: HEPiX Benchmarking Working Group co-chaired by Manfred Alef (KIT), Domenico Giordano (CERN) and Michele Michelotto (INFN Padua)
  - Several reports to GDB, HEPiX, WLCG MB
    - Domenico's report to MB 26-May-2020:
      https://indico.cern.ch/event/917098/contributions/3855129/attachments/2045174/3426154/WLCG-MB-26-05-2020-giordano.pdf
    - Domenico's report to HEPiX 13-Oct-2020:
      https://indico.cern.ch/event/898285/contributions/4034096/attachments/2121862/3571531/HEPiX-Workshop-13-10-2020-giordano.pdf
  - Without fixing the details of how to use it for a given purpose

- With respect to 2008...2009, landscape has changed completely, which implies new challenges and new opportunities

Future HEP Benchmarking (2)

- Result of the HEPiX benchmarking working group:
  - "HEP Benchmark Suite": Framework for running containerised benchmarking workloads
    - Automatises benchmark runs
    - Ensures structured delivery and storage of results
  - "HEP Workloads": Collection of (mostly) HEP workloads for which it is desirable to obtain performance information
    - Rather dynamic – add improved workloads, new compilers/flags/OS, …
    - Sometimes called "the matrix" or "the basket"
  - "HEP Score": Single number based on a defined, stable combination of defined, stable reference workloads; sample implementation ("HEP-SCORE20") using various LHC experiment workloads (from Run 2)
   - Good consistency with HEP-SPEC06

HEP-SCORE Deployment Task Force

- WLCG Management board discussed and decided to launch a task force

- Started in November 2020, bi-weekly meetings since then

- Membership:
  - Experts on pledge etc. process / procurements
  - Experiment experts
    - Four LHC experiments
    - Belle 2, DUNE, LIGO/Advanced VIRGO(/KAGRA), JUNO/BES III etc.
  - Site experts
  - Some MB members

Task Force Members

Name	Function	Name	Function
Tommaso Boccali	CMS	Jeff Templon	Nikhef
Simone Campana	WLCG	Andrea Valassi	LHCb
Domenico Giordano	Benchmarking WG	Ian Collier	STFC-RAL; APEL team
Michel Jouvin	Tier-2s	Gonzalo Merino	PIC
Walter Lampl	ATLAS	Fazhi Qi	JUNO, BES etc.
Andrew McNab	DUNE	Oxana Smirnova	NDGF
Helge Meinhard	WLCG and Convener	Tony Wong	US Tier-1s
Bernd Panzer	CERN	Josh Willis	LIGO/Adv. VIRGO
Stefano Piano	ALICE	Manfred Alef	KIT and Benchmarking WG
Randy Sobie	Belle 2	Alastair Dewhurst	STFC-RAL
Andrew Melo	Vanderbilt U; US CMS T2		
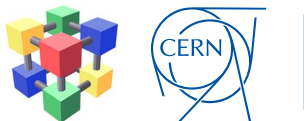
Task Force Topics

- Compute facilities at WLCG sites
  - Still very much x86 dominated (mostly Intel, some AMD)

- Compute facilities used (quasi-)opportunistically
  - Various processors (x86, POWER, ARM), e.g. at HPC sites
  - Various GPUs in various relations with CPUs, e.g. at HPC sites
  - May see some of this soon at WLCG sites, too – even as part of the pledges
  - More may be coming, e.g. FPGAs

Discussions and Agreements So Far

- Start with CPU benchmarking on x86-based systems; look at other CPUs and/or GPUs later
  - Aim: single benchmark with a stable definition for at least (a typical CPU server lifetime cycle | a complete LHC machine cycle)

- Framework by benchmarking WG is very attractive
  - Use it to record behaviour of (wide) selection of workloads (not limited to HEP) across machines

- No "final" candidate for new benchmark yet
  - Intense discussions whether HEP-SCORE20 is a starting point, and whether HEP workloads should be used at all
  - Decision deferred to after we have got hold of actual experiment workloads (LHC: Run 3?) and have collected data of selection of workloads (see above)

Task Force Activities

- Done
  - Reports to WLCG GDB and WLCG MB
    - MB endorsed
  - Link with accounting (WLCG accounting task force, APEL team) about changes to apply to accommodate new benchmark

- Being done
  - Round of status reports on workloads (most experiments done, DUNE and non-experiment workloads to come)

- Next steps
  - Implement workloads, once stable, into HEP Benchmarking Suite (the framework) to form HEP Workloads for data collection
  - Establish list of configurations to run HEP Workloads (the collection) on
  - Once ready, run on a variety of CPU server configurations to obtain behaviour of all workloads
  - Define representative sample of workloads for the final benchmark (HEP-SCORE2x)
    - Including relative weights
  - Propose steps for transition from HEP-SPEC06 to HEP-SCORE2x

Interested? How Can You Help?

- Integrate proposed workloads as containers into HEP Benchmark Suite

- Provide sample machines of typical worker node configurations

  – The more variety we have, the better

- Once ready, run HEP Workloads on your sample machines

(Personal) Conclusions

- Defining a HEP-SPEC 06 successor is a complex, ambitious task
  - Will take many months rather than weeks
  - … and then we will have to tackle non-x86 CPUs and GPUs
- Quite a number of different views on how exactly this should be done
  - Good representation in the task force
- Nonetheless, discussions in a constructive and collaborative spirit
- Steady progress

- Thanks to all task force members!

Questions? Comments?

?