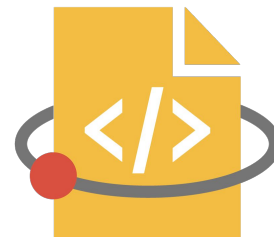# Software Update

Graeme Stewart (EP-SFT) and Liz Sexton-Kennedy (FNAL)

2020-11-17

Google Summer of Code & Season of Docs

- GSoC
  - We were awarded 36 slots from Goole
  - 34 students were successful in their projects
  - Google have announced that the coding time awarded in GSoC 2021 will be about *half* of previous years, which is not good

- Season of Docs
  - We have 4 projects in Season of Docs this year
    - AllPIX
    - Rucio
    - ROOT (x2 general + Python documentation)

CERN-HSF GSoC participation
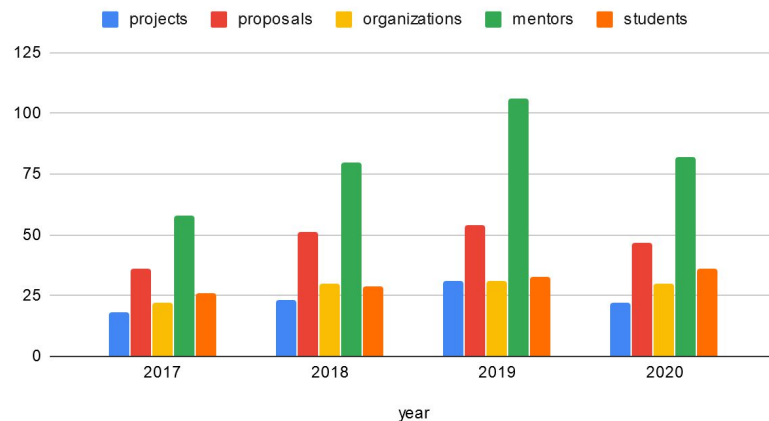
Event Generators

- Detailed presentation and discussion of the issues and progress in this area took place during the September LHCC meeting
- Dedicated session at the HSF-WLCG Workshop this week, focusing on two of the high-priority areas discussed in September
  - Reducing the impact of negative weights
    - MC@NLO-Delta (new theoretical calculation)
    - Neural resampler (ML approximation)
  - Progress on porting generator workflows to GPUs
    - MadGraph5_aMC@NLO
    - PDF Flow and VEGAS Flow
- Revised version of the expanded paper on event generation challenges for HL-LHC has been submitted to Computing and Software for Big Science
  - Referee comments addressed, arXiv:2004.13687

Geant4

- [Geant4 10.7-beta](#) preview-release deployed in July 2020, many *performance* and *physics improvements*
  - Integrated first version of Tasking system based on PTL or Intel/TBB
    - Paving the way for future sub-event level parallelism extensions
  - Extended configuration system to support modular builds
    - Integrated and enabled configuration of the built-in Python module
  - Enhancements to diagnostics system for field transport
  - Optimization of code handling internal data structures
    - Benchmarking ~5% speedup results for CMS geometry out of hit creation
  - New EM model for polarized gamma elastic scattering
  - New Coulomb scattering model for e+/e- based on very accurate Differential Cross-Sections using Dirac Partial Wave Analysis
  - New thermal model of positronium decay to gammas
  - Improved treatment of anti-baryon interactions in the Quark-Gluon-String (QGS) model
  - Revised maximum energy of applicability of elastic and inelastic cross-sections for ions and anti-ions nuclear interactions

- Bound with new release [VecGeom v1.1.7](#)
  - Revision of the build system and bug fixes
  - Enhanced navigation system with indexing and GPU adaptation for ray-tracing

- New patch release [10.6.p03](#) deployed in November for the current production Geant4 10.6

Geant4

- [Virtual] Geant4 Collaboration meeting, on 21-25 September
  - https://indico.cern.ch/event/942142/overview
  - Reviewed developments planned for release 10.7 in December
  - Identified key developments for inclusion in a major release next year
  - Collected requirements from users' community and LHC experiments
- CERN technical training [virtual] course (advanced) , 28 September – 2 October
  - https://indico.cern.ch/e/geant4_advanced_course_2020
  - Full attendance, limited to 45 attendees
- Review of Geant4 physics validation at LPCC Detector Simulation Workshop
  - LHC experiments + CALICE: https://indico.cern.ch/event/964486/timetable/
  - Valuable input for physics development and level of precision for detailed simulation
- Geant4 co-organising the detector simulation session at the HSF workshop

HSF Detector Simulation and R&D Topics

- Review of R&D projects for detector simulation
  - We identify a clear trend towards GPUs
  - We are consolidating efforts into one coherent R&D project to understand and to address the challenges related to running simulation code on GPUs
- Work has started at CERN on a *GPU prototype* for electromagnetic physics in calorimeters, **AdePT** (EP-SFT, IT-SC-RC, UK ExCALIBER HEP); DOE funding a similar project, **Celeritas**
- Meetings before summer break focused on this topic
  - Lightning talks on R&D projects for simulation on GPUs, including explorations in the context of the LHC experiments
  - Medical physics experience: hGATE and GGEMS[1], latter with physics effects from Geant4
- Simulation session at upcoming HSF/WLCG workshop
  - Review talks Geant4, ML-based fast simulations and experiment [technical] requirements
  - GPUs R&D more complex and/or advanced projects (AdePT, Celeritas)
- Next topical meeting on Deep Learning fast simulation
  - Validation strategies
  - Integration in experiments' frameworks

[1] **G**PU **Ge**ant4-based **M**onte Carlo **S**imulation

Reconstruction and Software Triggers

- Public GitHub organization for common trigger & reco code
  - Currently hosting code for tracking with graph ML Exa.trkX

- 19/11: joint meeting with Long-Lived Particle Community on use of reco software in non-LHC experiments (with a focus on long-lived particles/forward physics)
  - Many of those experiments share hardware/software with LHC experiments, showcase return-on-investment on common software development
  - Small experiments attending this workshop may find common software to use / take inspiration from
  - Confirmed: CODEX-B, FASER, LDMX; Future: MATHUSLA, HPS, LHCf, MilliQan, MoEDal
- Planning December meeting: summary of FastML workshop
 - Topic: strategies to accelerate trigger and reconstruction with machine learning
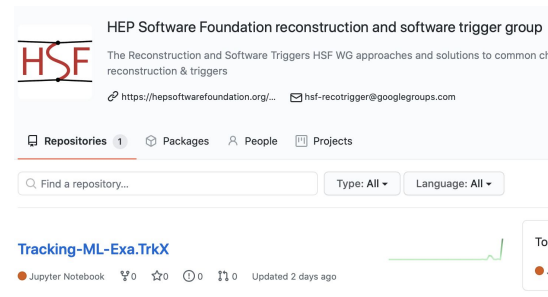
ROOT

- Advances especially in new HL-LHC data format ("RNTuple" - faster and more compact), declarative analysis ("RDataFrame"), statistical modeling ("RooFit"). Released v6.22 in June, e.g. with new PyROOT after two years of work. Ongoing llvm upgrade since March to provide C++17 support for EDM as requested by experiments.
  - Presentations at ICHEP (ROOT and RooFit), CEPC, EIROForum
  - Very well received offline + remote course with CMS; ATLAS is employing trainers from ROOT's "train the trainers" event
- Very smooth progress and productive team, despite COVID; adjusting well to ongoing changes. Major difficulty is integration of newcomers!
- Shortage of effort in the team is now becoming dramatic:
  - 3 staff FTEs down for many months, such as computing coordinator "on loan"
  - External senior developer's end-of-contribution, mostly ineffective summer student year, two interns without replacement (COVID hampered these contributions); 1/2 junior fellow since Nov, but in no way can make up for the above reductions
  - *Progress on hold in many areas until FTE back at reasonable level*
    - Impossible to make progress with HL-LHC generational hand-over, for code or team - focusing on maintenance and support
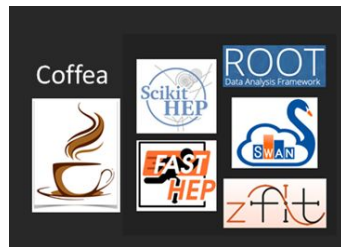- Discussion of plan of work for 2021 started

HSF Analysis WG

- Participated in Snowmass Computing Frontier kick-off

- LHCC review priorities set the themes for 2021 activities, planning in progress
  - Analysis resource usage & computing model
  - Declarative analysis -- interfaces & backends
  - Metadata handling for analysis -- calibration & book-keeping
  - Quality assurance: job curation & code testing

- Aim to increase cooperation with other WGs
  - Training: improve resources for new analysts, introduce tools for effective SW development
  - PyHEP: envision python stack for analysis?

PyHEP WG
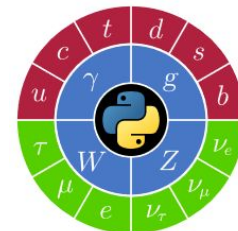
- 3rd PyHEP workshop, PyHEP 2020, held on July 13-17
  - Was to be co-located with SciPy 2020 in Austin TX, but both became virtual due to COVID-19
  - PyHEP 2020 agenda organised in 2 time zones to accommodate Asia, Europe and Americas
  - Remarkable level of interest - *we limited at 1000 registrations!*
  - 2 keynote talks and ~30 hands-on tutorials and "notebook-talks"
  - Various tools and procedures tried, with very positive feedback from participants
    - Topical Slack channels for communication, Slido for after-talk Q&A sessions, notebook talks launchable online with Binder (dedicated resources), recordings *captioned* and uploaded to dedicated YouTube playlist, all presented materials given a DOI via Zenodo
- Topical meetings being planned for 2021
  - Interest from a growing community, with several experiment-agnostic projects

  - https://github.com/CoffeaTeam
  - https://github.com/FAST-HEP
  - https://github.com/root-project/
  - https://scikit-hep.org/
  - https://github.com/zfit

Software Process and Infrastructure - SPI

- Regular delivery of software stacks (LCG releases) for ATLAS, LHCb, SWAN and BE, paced following the major releases of ROOT and adapted to the needs of each specific experiment
  - Current release LCG_98 with ROOT 6.22
- Includes latest versions of Monte Carlo generators in agreement with authors and experiments
- Building experimental LCG stacks for special cases, such as the ARMv8 architecture, OS X or support for GPUs (currently specifically CUDA)
- Software support for SME experiments (NA61/SHINE, NA62) and for the Continuous Integration of projects (DD4hep, Gaudi, tkLayout, Allen, …)
- Provision of latest compilers (e.g. Gcc, Clang, CUDA) and related tools (e.g. GNU Binutils, CMake) for the community through CernVM-FS
- Continuous support of LCG stack users (JIRA/Helpdesk tickets)

SPI - R&D and Remarks

- Investigate the use of the Spack package manager to build and deploy the stacks
    - Prototype work based on the LCG stacks
    - Concrete working example for the Key4hep experimental stack, supported by EP R&D project
- Investigate the use of Kubernetes to orchestrate containers builds
    - Includes the use of CernVM-FS conveyors to streamline publication
- The interest in LCG builds is increasing, more and more SME experiments and projects are interested in it.
    - This is partially due to the attempted AFS phase-out/ramp-down, but it is going to stay.
- The LGC stack is dynamically evolving. During 2020 ~75 new packages have been added to the stack, ~25 removed, with an overall growth of ~10% .

CernVM

- Established /cvmfs/unpacked.cern.ch as a HEP community-wide container hub
  - >500 container images, >65M files
  - Efficient distribution at WLCG scale, 10x-50x more efficient than container distribution with off-the-shelf tools
  - Ready-to-use with singularity container engine
- Proof-of-concept integration of popular container runtimes podman (new Red Hat default) and containerd/kubernetes
- CernVM-FS 2.7 feature release plus handful of patch releases
  - Code modernization and adoption of recent Linux capabilities necessary for better container support and faster publishing and reading
- Future proofing LHC experiment repository publishing service: IT-ST together with EP-SFT migrated >20 LHC experiment and dependent repository publishers from EL6 & block storage to EL7 & faster Ceph S3 backend storage

CernVM - Upcoming Items

- Preparation of the CernVM virtual workshop 1-3 February 2021
  - Presence workshop at NIKHEF postponed to 2022
- Ongoing effort to meet higher publishing throughput demands of container and LHC nightly build repositories
  - >1M files per day and nightly build repository added and garbage-collected, upward trend
  - First production exploitation phase of distributed publishing service started with LHCb
- Software deployment on HPC remains challenging because there is fewer common ground among HPC sites than we are used to from the grid
  - Continuous effort on trust building, e.g. new collaborations with HPC sites
  - Requirement for more advanced CernVM-FS caching capabilities to cope with large HPC sites with diskless compute nodes
- Evolution of the CernVM appliance to become a common platform compatibility container
  - Addresses the container proliferation problem, especially for analysis (not every analysis container should need to bundle its own operating system layer)
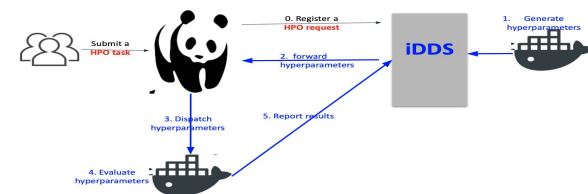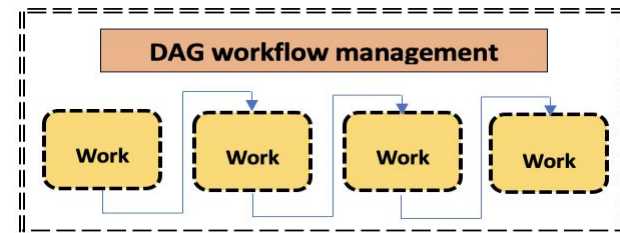
iDDS (intelligent Data Delivery Service)

- Allied HSF activity from IRIS-HEP/ATLAS with strong links to DOMA
- Data Carousel
  - In production since May; issues related to startup of operations from tape fixed
- HPO (Hyper Parameter Optimization)
  - Provide a fully-automated platform for hyper-parameter optimization on top of geographically distributed GPU resources on the grid, HPC, and clouds
  - Advertised to ATLAS ML users, not specific to ATLAS
- DAG based workflow management
  - High-level workflows specified by DAGs driving workload scheduling
  - Tested using new DOMA PanDA instance for Rubin Observatory (LSST) exercise
    - Cascade of chains for multi-step processing with thousands of jobs per step
    - Release jobs incrementally for different steps to avoid long waiting
- Near-term plans
  - Improvements of user experience
    - Client and CLI development and improvements, documentation and monitoring
  - Endeavour to support more use cases in multiple experiments
    - DAG based Active Learning is ongoing
    - Dynamic transformation and placement on demand
    - Fine-grained data transformation and delivery, such as Event Streaming Service
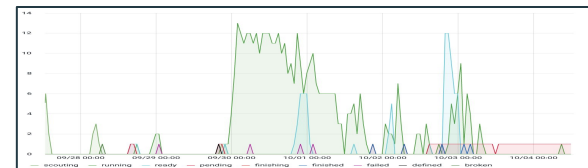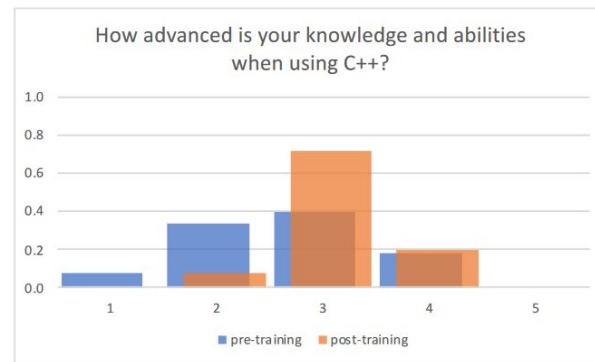

iDDS tasks accounting (by status)

iDDS HPO

iDDS DAG

Training

Monday	Tuesday	Wednesday	Thursday	Friday
Basics (syntax, pointers, references, compound types, objects ...)	**Object Orientation** (classes, inheritance, virtuality, operators, ...) **Constness**	**Tools** (git, compile chain, gdb, valgrind suite, ...)	**Core modern C++** (move semantic, templates, lambdas, STL, RAII, ...)	A la carte: **Concurrency, Python & C++, Expert C++** (poll in the subscription form)

- Many training events have been taking place
  - These have been virtual events, obviously, but that does have some advantages
    - Pipelines Training
    - Docker Training
    - GPU and ML Training
- C++ Training (co-organised with SIDIS)
  - Long discussed, it finally happened!
  - Ambitious course taught by Sebastien Ponce (LHCb) with 10 mentors (thanks!)
  - Hugely oversubscribed
    - 50 people accepted onto the course
    - 179 people on the waiting list, huge appetite for this kind of training
  - Outcomes were very positive
 - Will repeat in the New Year
- Challenge is to put all of this training on a **sustainable basis**, with suitable rewards for tutors and helpers
  - *We have a dedicated slot at the HSF Workshop on training to help to build up this training community*

How advanced is your knowledge and abilities when using C++?

pre-training   post-training

HSF WLCG Workshop

- As mentioned already we are having the second edition of our virtual workshop this week
- To focus more in depth we are running computing and software sessions in parallel
- We decided to have an open R&D session with abstract submission
  - This proved to be very popular (x2 over subscribed)
- 341 registrations
  - Which is our most popular workshop to date
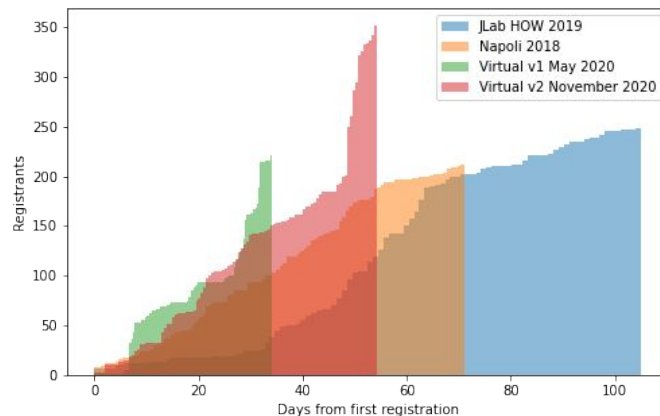
Compute Accelerator Forum

- We have noted in the software development community that there is a move away from pure CPU processing
  - This brings many opportunities, but many new challenges to development teams
- HSF, SIDIS and openlab identified the need for a forum to discuss *fundamental aspects* of programming on these new devices
  - Development resources
  - Abstraction mechanisms
  - Data layouts
  - Toolkits and performance optimisation
- Plan to hold <u>monthly meetings</u> on these topics…
  - October - Access to GPU resources at CERN and in WLCG
  - November - Abstraction mechanisms employed by ALICE and LHCb
  - December - Performance monitoring tools from Nvidia

Summary

- A range of activities going on, supporting current and future experimental programme, *key focus on what the experiments need for simulation, reconstruction and analysis*
- Technological landscape remains fluid
  - GPUs have many R&D projects bearing fruit (still tricky to use and do not solve all problems!)
  - Newer CPU architectures could claw back some of the HL-LHC gap? (ARM needs rebuild, AMD is ~free)
- Training and R&D often go hand in hand
  - Blends into an expert discussion on how best to use new tools
- HSF providing a general forum for improving and sharing ideas
  - New initiatives such as the Compute Accelerator Forum
- EP-SFT projects are critical for the experiments
  - The load of production support and R&D puts a lot of pressure on teams
  - Human resources often at quite a critical level and perturbations then have a serious impact