

Track 5 Highlights Sustainable and Collaborative Software Engineering

Giulio Eulisse, Eduardo Rodrigues, Liz Sexton-Kennedy

Topics

Sustainable Languages and Architectures

Sustainable CI and Build Infrastructure

Glance and Web based applications

Sustainable Analysis

MISC: MonteCarlos, Infrastructure and Simulation

Sustainable Frameworks

A total of 33 talks in 6 sessions.



Track 5

A Theme for this CHEP: Increasing adoption of open source tools outside of HEP & NP Theme of this Track: Sustainability through sharing of common tools



Sustainable Languages and Architectures

Interoperability between different languages and different hardware platforms was a key topic of this session.

Two nice talks about Julia being a sweet spot between C++ performance and Python ergonomics, including one with a concrete example implementing FastJet algorithms.

> "Julia is impressive, it's easy to work with and fast." -Graeme Stewart

Future of Programming



Larger values are better

HepScore numbers for ATLAS reconstruction

HepScore integration (II) - reconstruction scores, full node

- Node fully packed with n×4 threads
- \rightarrow Different number of available cores explains different scales !
- Orange/Brown/Blue: ARM flavours
- Green: Intel/AMD flavours
- Some hyperthreading or IO related differences
- Similar trends for Event generation and simulation
- See back-up for single workflow measurement

Which language would we have picked in 2013 if we had to choose from today's programming languages?





Sustainable Languages and Architectures

This session also had 2 talks about engaging and training the community:

User-Centered Design for the Electron-Ion Collider

Markus Diefenthaler

mdiefent@jlab.org

- Reality: On average, 78% of students' and postdocs' research time is devoted to software and computing.
- Goal: Enable active participation in physics analysis, regardless of career stage, beyond just students and postdocs.
- Solution via User-Centered Design: Engage community in development. Listen to users, then develop software.
- Started projects on User-Centered Design in ePIC.
- User-Centered Design is part of our software principles.





Scalability and Sustainability



Paradigm

- No standard curricula for HEP students exists
- Not all HEP students can attend university-offered software courses
- HEP students in many cases don't receive any programming training
- Students trained as physicists but asked to be data analysts

^ Lessons build on each other

A Enough verbosity for self-study

· demonstrate how to move an e

The skills we'll focus on



We need a unified, scalable, and sustainable software training framework powered by the entire community

HSF/IRIS-HEP is leading training efforts

Most training modules are website built from

Machine Learning on GPU < Complete video walkthroughs!

IRIS-HEP is generating a wealth of training materials in addition to organizing training events

rence Highlights: Track 5

Sustainable CI and Build Infrastructure

A common trend across the experiments is orchestration & containers using kubernetes or nomad.

Jenkins still heavily used.

Gitlab / Github actions new-ish way of defining builds.



ARCHITECTURE OVERVIEW



Glance and Web based applications

The Glance project is providing IT services specific to the needs of the experiments.

Sharing infrastructure allows improvements to propagate. We heard about - synchronization, search, and infrastructure improvements.

mouting in High Energy & Nuclear Phy



(•) ance presentations @ CHEP 2023: Other

The ALICE Glance Service Work system Jomar Junior de Souza Pereira | 08/05, 15h | Track 8

The ALICE Glance Membership System Jomar Junior de Souza Pereira | 09/05, 15h30 | Poster 13

Iterative and incremental development of the ATLAS Publication Tracking system Ana Clara Loureiro Cruz | 09/05, 14h | Track 5

Glance Search Interface Carlos Henrique Ferreira Brito Filho | 09/05, 14h30 | Track 5

The migration to a standardized architecture for developing systems on the Glance project Carlos Henrique Ferreira Brito Filho | 09/05, 14h45 | Track 5

> Contact gabriel.aleks@cern.ch

Used by 3 LHC experiments with 12 active developers supporting +20 web systems,

Sustainable Analysis

Coffea = user interface for columnar analysis:

- dask_awkward fundamentally changes how we can describe analysis
- dask_awkward based analyses, via dask task graphs, are rendered into a general, complete, declarative analysis description language (ADL)
- Represents the culmination of ~4 years of R&D

Law = Luigi Analysis Workflow, in Python

- Large Scale End-to-End Analysis Automation over Distributed Resources
- Designed to fully decouple these 3 aspects:







Sustainable Analysis

"Analysis of physics analysis"

Idea to learn from Git repositories. Study limited to CMS



- We can learn things that are useful for software library maintenance
 - user adoption of new versions
 - most common function-call patterns
 - decide if and when a feature can be deprecated
 - discover which libraries are being used together, maybe motivate integrations

LbMCSubmit:

- New flexible & scalable request submission system for LHCb simulation
- Based on specifications in YAML, with review, CI test, then full submission -

Simulation data quality in LHCb now integrated with DQ monitoring web-based tool



Fri 12th May Plenary Session - Conference Highlights: Track 5



sim-version: 09 name: Ds2KKpi inform:

– event-types:

- 2016 num-events: 100 000

- 23103005 - 23103006 data-types: - 2012

- firstname.surname@cern.ch

- auser

WG: Charm

MISC: MonteCarlos, Infrastructure and Simulation

- EPOS 4 General-purpose event generator for hadronic physics out since Oct. 2022
 - RIVET now integrated into the EPOS analysis system for effective testing and best parameters tuning (required in the validation of release cycles).
- GEMC Can we define a geant4 simulation in its entirety from a database?
- LinearColliderIO Event Data Model turns 20 !
 - Long-term goal was to encourage interoperability of worldwide LC simulation/reconstruction/analysis programs
 - LC community moving to key4hep and EDM4hep in an adiabatic way (LCIO<->EDM4hep conversion available)
- Darshan is a lightweight I/O characterization tool that captures concise views and entire traces (DXT) of applications' I/O behavior
 - Enhanced to also work in non-MPI contexts (HEP use cases)
 - Being used by ATLAS/CMS to study I/O workflows





Fri 12th May Plenary Session - Conference Highlights: Track 5

GEMC: turnkey database-driven MC simulations program

MISC: MonteCarlos, Infrastructure and Simulation

- Gauss-on-Gaussino = new version of Gauss (Gauss = the LHCb simulation framework)
 based on Gaussino with LHCb-specific additions
 - Gaussino = new core simulation framework, created by extracting experiment-independent components from Gauss

Gaussino can be used (1) as the core simulation toolkit for detector frameworks in HEP or (2) in a standalone mode as an ideal test bed for early detector studies. ML inference in fast simulations

2 possibilities for ML inference

- 1 internally in Geant4,
- in a Gaudi algorithm after the simulation algorithm takes place,

First benchmarks

- in a close collaboration with Geant4/ML4Sim,
 - compatible with the CaloChallenge setup,
 - 🗲 generic cylindrical calorimeter setup,
 - variational autoencoder for photons,
 - Gaussino's interface to pyTorch C++ API,



Key4hep: Turnkey Software Stack



Sustainable Frameworks

Key4hep is a turnkey system providing components & building blocks for many efforts in the future collider world.

Meld: Exploring the feasibility of a framework-less framework - for DUNE Functional programming approach to Frameworks. Structured software stack integrating individual packages towards a complete data processing framework for HEP experiments

 Reduce overhead for adopting projects by sharing common components

Functionality-complete: plenty of examples for simulation and

Easy to use for librarians, developers, users

reconstruction of detectors



Preserve and adapt existing functionality into the stack, e.g., from iLCSoft, FCCSW, CEPCSW

Main ingredients

Event Data Model (podio/EDM4hep) (see **talk by T. Madlener**), Geometry Information (DD4hep), Processing Framework (Gaudi), Package manager (Spack)

ess Report on Integrations - CHEP 2023, May 11, 202





Sustainable Frameworks

The Open Data Detector Project is trying to establish the next generation detector modeling tools for algorithmic and computing R&D.

Tracking/Calo challenges w/in the Key4Hep ecosystem.

Open-sourcing and Publication of the Belle 2 software stack.



Impact?

- * No negative consequences observed so far
- Software citations
- Google Summer of Code
- Job application support?





Fri 12th May Plenary Session - Conference Highlights: Track 5

Thomas Kuhr (LMU)

A Big Thank you to the CHEP organizers and all the participants for making this CHEP a truly enjoyable experience; it's great to be back together again !



