Software Training in HEP

Sudhir Malik
University of Puerto Rico Mayaguez
(on behalf of HSF/IRIS-HEP training group and all contributors to the training)

Co-authors:
Samuel Ross Meehan (CERN)
Kilian Lieret (Ludwig Maximilian University Munich)
Meirin Oan Evans (University of Sussex (GB))
Michel Hernandez Villanueva (University of Mississippi)
Daniel S. Katz (University of Illinois)
Graeme A Stewart (CERN)
Peter Elmer (Princeton University (US))
And many more at: https://hepsoftwarefoundation.org/training/community.html
Software a key to HEP success

- Solving software challenges integral to the success of current and future HEP experiments (HL-LHC, DUNE, etc.)

- Software and computing systems are key subsystems of our experiments, involve significant budget

- Maximizing science from the hardware investments increasingly relies critically on software

- Software skills essential for a successful HEP physicist, and for career evolution for people trained inside HEP, seeking career in industry
Scientific Collaborations are big and growing

- Current examples (estimate stats)
  - **BELLE II** - 1200 collaborators/121 institutes/26 countries
  - **CMS** - 4000 collaborators/200 institutes/50 countries
  - **ATLAS** - 3000 collaborators/174 institutes/38 countries
  - **LHCb** - 1200 collaborators/76 institutes/16 countries
  - **ALICE** - 1000 collaborators/100 institutes/30 countries
  - **DUNE** - 1000 collaborators/180 institutions/30 countries
  - **LIGO** - 1200 collaborators/100 institutions/18 countries

- Past
  - **DZero** - 540 collaborators/90 institutions /18 countries
  - **CDF** - 600 collaborators, 30 institutions/12 countries
HEP software ecosystem

Software and Physics analysis are intertwined

Physics Event Generators

Detector Simulation

Trigger, Event Reconstruction

Data Analysis, Interpretation, Simulation

Visualization

Data Processing Frameworks

Machine Learning

Lots of challenges

Data, Software, Analysis Preservation

Security

Software Development

Facilities, Distributed Computing

Data Management Organisation, Access
HEP Paradigm

● Paradigm for HEP users
  ○ Knowledge of complex computing and physics analysis tools intertwined
  ○ Software challenges wrt to data rates, processing and analysis
● Long life span of the experiment ~ 30 years
● Enormous data rate
● Most users not resident at host laboratory
  ○ Financial and logistic constraints to be at host lab (e.g. CERN)
● Highly distributed environment for
  ○ Computing (Grid)
  ○ Physics analysis
● Physics/Computing Support
  ○ Should reach every user wherever they may be
  ○ Should be taken up in organized and central way
Training Challenge

● Training is a prerequisite to meet data and software challenges

● Funding agencies and institutions may not have the same priority for software training and education as for building/operating detectors, physics analysis, etc.

● Training activities are severely undervalued in making career steps

● Individual universities do not uniformly provide training today prior to a Ph.D. student beginning their research career

● Volunteers can usually only dedicate their time in specific career phases as a side “hobby” project

● Training materials are a moving target as technology evolves

● Separating “Experiment” specifics (e.g. computing environments or dedicated software) from HEP wide “common” usable material is important, but doesn’t always happen

● Are training materials a common good or an individual product? Even if individuals do want to contribute to a common good, how do they do so?
Training Vision

- Provide training in the computing skills to produce high quality sustainable software, solve HEP challenges, software-trained workforce

- Train every new HEP entrant in the related Software and tools

- Build community for scalability and Sustainability

- Training style - Hands-on, Student-centric, Experiment Agnostic, Reuseable, Open and Accessible

- Broader impact - reach out to STEM aspirants in high schools via their teachers and involve students early on
Training Organisation

- Led by HSF training group, established ~3 years ago
- 3 co-convenors, engaging with educators from different collaborations
- Strong partnership with IRIS-HEP, FIRST-HEP and the Carpentries
- Prepares training material and coordinates activities for the common good
- Strong community of instructors and participants, feeling of community ownership
- Focuses on common software material across HEP, ranges from basic core software skills needed by everyone to advanced training required by specialists in software and computing
Curriculum

- The pilot phase of training events was based on an initial survey across HEP community in 2019.
- This survey and experience and feedback gathered at the events lead course structure into a full curriculum.
- Guidelines for the development of the modules and the procedure for training events are formalized.
- Each training module is independent from the others, students can prioritize certain skills before others.
- All software material is open source.
Accomplishments

● **Software modules**
  ○ Basic software curriculum
    ■ Introductory software training curriculum serving all HEP entrants
  ○ Intermediate modules, some specific to HEP
  ○ All modules are open source

● **Training events** (last 2 years)
  ○ 13 events, 1500 participants
  ○ In-person, Online (Covid Impact)
  ○ Out of these, 5 are Basic Curriculum
    ■ 400 attendees

● 120 instructors involved
Trainings to Date

- **2020**
  - "The Awesome Workshop" (30 participants / 10 educators)
  - LBNL ATLAS Software Bootcamp (40 participants / 8 educators)
  - Python & Stuff @ FNAL (25 participants / 5 educators)
  - CICD with GitLab/Pipelines [virtual] (250 participants / 15 educators)
  - Containerization with Docker [virtual] (173 participants / 15 educators)
  - US+CA ATLAS Computing Bootcamp [virtual] (50 participants / 15 educators)

- **2021**
  - Sebastian and Stefan’s C++ [virtual] (50 participants / 12 educators)
  - Machine Learning + GPUs [virtual] (40 participants / 7 educators)
  - C++ Training
  - CICD with GitHub/Actions (lesson content)

You want to have a high impact and advance HEP? - Training might be your most effective choice!

- 2019
- 2020
- 2021
In person training

- **Attendance**: few *dozen*

- **Advantage**
  - Active/efficient engagement of participants
  - Professional networking and additional “events”

- **Limitations**
  - Travel costs (education should not be exclusive)
  - Long lead time for planning logistics
    - Related to travel/room booking
  - Requires participant “sacrifice”

- **Important things**
  - Room setup is crucial
    - Two projects/screens
    - Not an auditorium
    - Ample power

- **Suggested Ratio of Participant**: Educator <= 5
  - This is *essential* to allow for the “hands on” aspect of the workshop to be successful

- **Large time commitment** on behalf of the educators
  - Can’t just “do your talk” and then leave
Virtual training

- Pivot to remote training due to COVID
  - Adapted quickly
  - 7 online training workshops (last 12 months)
- Attendance: few hundred
- Positives
  - Broader reach, more participants: >100 registrants
  - No travel costs → critical for some supervisors
  - Easier logistics, easy to reach all timezones
  - Materials are recorded and archived (videos)
- Limitations
  - Active/meaningful interactions
  - Mentors in different time zones
  - Keep everyone engaged
- Important things
  - Clearly-defined roles for instructors
  - Effective chat application is essential
    - mattermost/discord/slack
Training Works !!

- We do our best to diligently collect before/after data via surveys
  - Pre-survey
    - Demographics
    - How much do you know?
  - Post-survey
    - How much do you now know?
    - What can we do better next time?
  - Would like to have further out “follow up” surveys (takes more work …)
- Self-reported learning *does* happen!
Impact and diversity

GitHub CI/CD Training Example (Feb 2021)

Experimental Collaboration
- ALICE
- ATLAS
- CMS
- LHCb
- Neutrino Experiment (DUNE, ProtoDun…)
- Smaller Experiments (FASER, NA62/6…)
- Direct Dark Matter (LUX, Xenon, etc.)
- Theory Community

Gender
- Male
- Female
- Non-binary
- Prefer not to answer

Academic Level
- Bachelors Student (Undergraduate)
- Masters Student
- PhD Student (1st or 2nd year)
- PhD Student (3rd year or more senior)
- Post-doctoral Researcher
- (Permanent) Faculty/Staff Scientist
- Staff, computing operations
- Staff at CERN

Location
- Europe/Africa
- South/North America
- Asia/Australia
Lessons Learned

- Advantages and limitations of in-person and virtual trainings
- Build a community around training
  - Incentivize and compensate instructors
  - Core team to support the training mission
- Scale up training (new formats)
  - Core format: we organize and teach
    - In-person, Online
    - Fund instructors to travel and teach
    - To scale up, need to expand to other formats
  - DIY (Do-it-yourself)
    - Minimal help from us, basically organise yourself, using training material (no expense involved)
    - In-person, Online
  - Asynchronous (Anytime/Anywhere)
    - Coursera-type, small professional videos (~10 mins.), Q&A assessment
    - Use current material to extend training to this format
Community

- Active community members to support training
- Time dedicated on voluntary basis, great dedication and enthusiasm
- Tutors come from different HEP collaborations
  - This diversity adds great value to the training
  - Brings flavor of experience from a different computing environment
  - Common goal to create, teach, and sustain a common set of skills across
- Prepare for careers in software, strengthen job profile and enhance chances of employability in industry
- Profile of each tutor that contributes to the training on HSF page
  - Public proof of their capability, skills and contribution
Sustainability

● **Training model**
  ○ Training without borders
  ○ Strong community would lead to minimal human resources needed to keep training infrastructure running
  ○ Long-term financial support

● **Build regional and local capacity**
  ○ Empower HEP communities
    ■ Local mentorship and leadership
    ■ Engage more HEP labs and universities

● **Opportunities to grow professionally**
  ○ Career paths, strengthen job profile

● **Equity, diversity, inclusion and accessibility**
  ○ Participation across HEP communities, under-resourced, underrepresented institutions, communities in different geographical regions
    ■ Serve as a role model
    ■ Open source is a step in this direction
Broader Impacts

● Organized 6 outreach events
  ○ Programming
  ○ HEP data preview
  ○ CMS Open Data
  ○ Machine Learning basics
  ○ Machine Hackathon
  ○ Events in-person and online

● Future Outreach
  ○ More events per year in HEP-related communities
    ■ Keep in mind: Teachers available only at end of semesters
  ○ Develop short video modules for teachers and students to learn software anytime
  ○ Supplement workshops (in-person or online)
  ○ More engagement with Quarknet
Upcoming Events

- Basic curriculum training
  - July, September, December 2021
- Brainstorming sessions
  - June and November 2021
  - April 2021
- Matplotlib for HEP (October 2021)
- Modern C++ (September 2021)
- Introduction to Singularity (November 2021)
- HEP data analysis
  - Advanced analysis tools in HEP
  - February 2022
Summary

● Software training is making a difference
● Organized several training events in-person and virtual
● Learnt valuable lessons
● Virtual training has increased impact
● Developed basic software curriculum modules
● Intermediate and advanced level modules are populated
● Next step is to scale up training activities
● Synergy among HEP experiments (including neutrino) and Nuclear Physics community exists
● Preparing trained workforce pipeline
● Broader Impacts, inclusiveness and diversity are integral part
Training Information

- **Training events**: [https://indico.cern.ch/category/11386/](https://indico.cern.ch/category/11386/)
- **Material**: All the training modules developed so far resides: [https://hepsoftwarefoundation.org/training/curriculum.html](https://hepsoftwarefoundation.org/training/curriculum.html)
- **Community**: Our training community is listed here: [https://hepsoftwarefoundation.org/training/community.html](https://hepsoftwarefoundation.org/training/community.html)
- **Procedure**: how to request and organize a training: [https://hepsoftwarefoundation.org/training/howto-event.html](https://hepsoftwarefoundation.org/training/howto-event.html)
- **Funding**: Funding for training events is provided by the IRIS-HEP/FIRST-HEP
- **Blueprint**: First blueprint on training [https://indico.cern.ch/event/889665/](https://indico.cern.ch/event/889665/)
- **Videos**: [https://www.youtube.com/c/HEPSoftwareFoundation/videos](https://www.youtube.com/c/HEPSoftwareFoundation/videos)
Acknowledgments

Our community is growing and credit goes to many individuals, especial thanks to:

- The US National Science Foundation through grants OAC-1829707 and OAC-1829729 (FIRST-HEP), Cooperative Agreement OAC-1836650 (IRIS-HEP)

- Hosts and partners of training events mentioned in the talk:
  - CERN
  - Fermi National Accelerator Laboratory
  - Argonne National Laboratory
  - Lawrence Berkeley National Laboratory
  - The Carpentries
  - The US-ATLAS
  - Princeton University
  - University of Manchester
  - University of Puerto Rico at Mayaguez