FIRST-HEP training program

Sudhir Malik - Univ. of Puerto Rico Mayaguez

2019 Joint HSF/OSG/WLCG Workshop HOW2019 - Jefferson Lab (19-23 March, 2019)

FIRST-HEP

- Stands for - *Framework for Integrated Research Software Training in High Energy Physics*
- Collaborative proposal (Princeton+UPRM)
  - http://first-hep.org
- **Funding from NSF**
  - Training Workshops/Participant Support/Brain Storming Session

OAC-1829707 OAC-1829729

IRIS-HEP

- "*Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)*" - http://iris-hep.org
  - Sharing vision and co-supporting/organizing FIRST-HEP activities
  - Funding for Training Activities/Participant Support
  - Fellowships to work on Software and Computing Projects
  - Job Opportunities

OAC-1836650

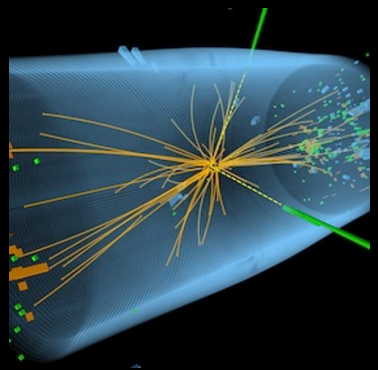
HSF
HEP Software Foundation

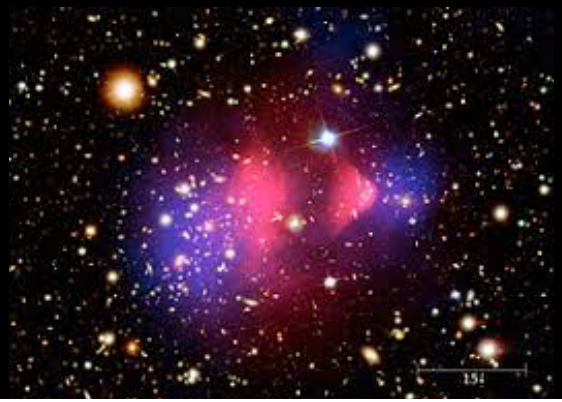
International Vision

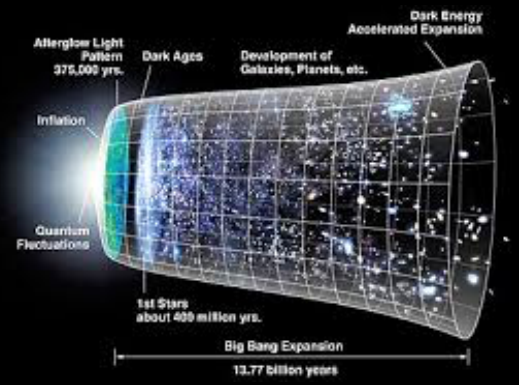
- The HEP Software Foundation (HSF)
  - formed in 2015 facilitates cooperation and common efforts in High Energy Physics software and computing internationally
  - strong training component to build HEP domain knowledge + advanced software skills+ strong connections to other related disciplines
  - https://hepsoftwarefoundation.org

Science Drivers

- Use the Higgs boson as a new tool for discovery
- Pursue the physics associated with neutrino mass
- Identify the new physics of dark matter
- Understand cosmic acceleration: dark matter and inflation
- Explore the unknown: new particles, interactions, and physical principles
- Bases of operation of many current experiments as well as the design of the large, next-generation, facilities in HEP HL-LHC, LBNF at Fermilab, Super KEK-B and experiments - ALICE, ATLAS, CMS, LHCb, DUNE, Belle- II

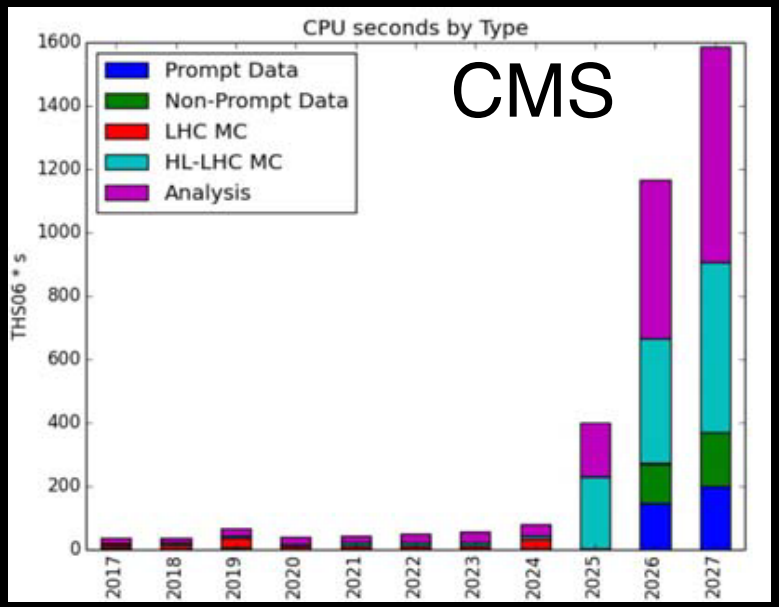


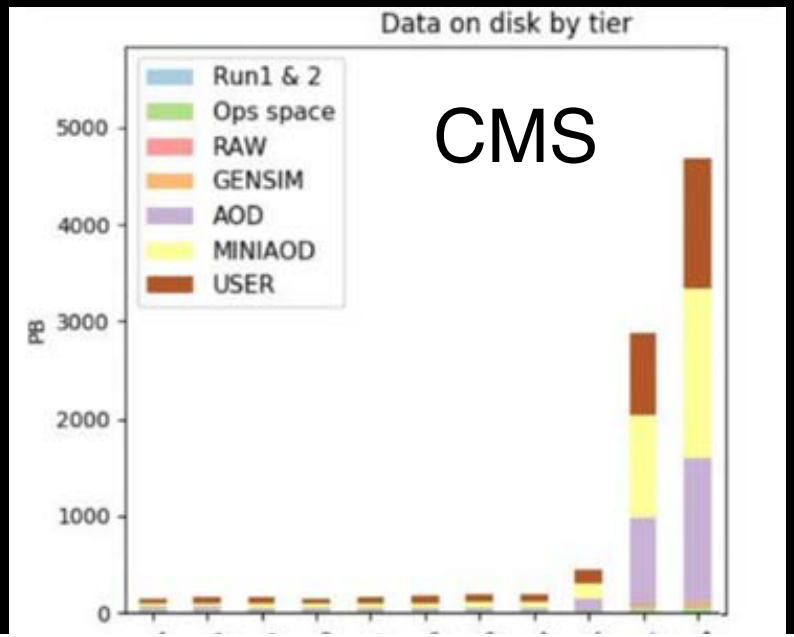




Computing Challenges in HEP

- Large data-intensive HEP experiments rely on
  - Significant data storage
  - High throughput computing
  - LHC Experiments - ~170 computing centers, nearly an exabyte of disk and tape storage and 750,000 CPU cores
  - HL-LHC
    - 100 billion proton-proton collisions per year
    - 10x or more computing needs, 100 times the data of (unto 2030s)
    - Other HEP facilities are planning similar increases in data volume

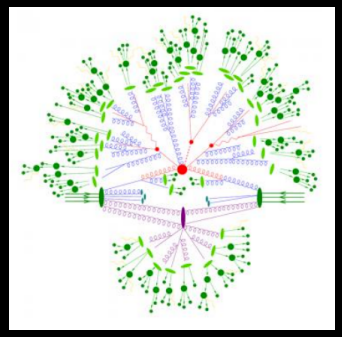




Questions guiding software & computing R&D

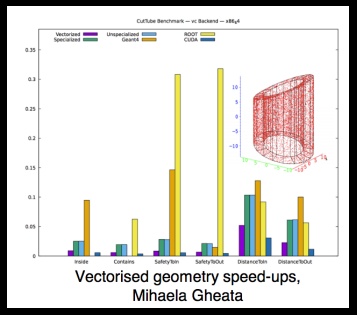
- **Physics:** Will efforts in this area enable new approaches to computing and software that maximize, and potentially radically extend, the physics reach of the detectors?
- **Resources/Cost:** Will efforts in this area lead to improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies, that allow the experiments to maximize their physics reach within New tools requires investing in training
- **Sustainability:** Will efforts in this area significantly improve the long term sustainability of the software through the lifetime of the HL-LHC?

HEP software ecosystem



Physics Event Generators

Data, Software,
Analysis Preservation

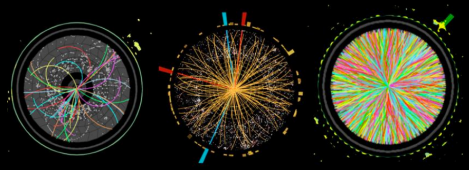


Detector Simulation

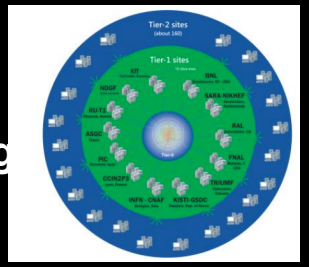
Security

Software Development

Trigger,
Event Reconstruction



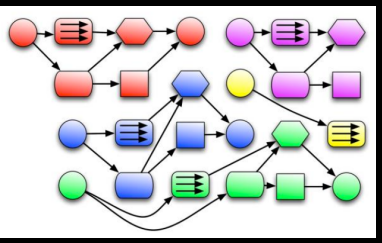
Facilities,
Distributed Computing



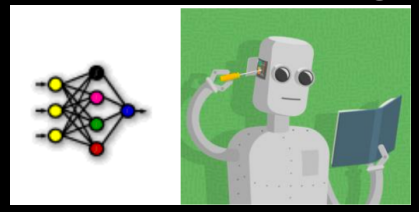
Data Analysis
Interpretation

Visualization

Data Processing
Frameworks



Machine Learning



Data Management
Organisation Access

Sustainability

- Successful evolution of this ecosystem to meet the challenges, requires new tools and a workforce HEP domain knowledge and advanced software skills
- Investment in SW critical to match HL-LHC requirements of “flat budget” scenario
- Investing in training leads to preservation and propagation of knowledge
- Investing software skills is not only important to actually build the requisite software infrastructure, but will also change community norms, create role models and promote career paths
- Computation is a central element of 21st century science, and clearer career paths will provide a virtuous cycle of feedback to enhance the vibrancy of the training and workforce development activities

Current Training Practices

- Training activities today are fragmented and partially redundant
- Each project (experiment, laboratory, etc.) is left to reinvent most aspects of research software training from scratch, the result is duplicative and often incomplete
- Most training activities are carried out “locally”, with specific objectives in the
- The modest effort devoted to training is not always positioned for maximum impact.
- The resulting activities are also quite difficult to sustain over time
- They are too often critically dependent on specific individuals whose careers evolve
- The effort to keep training materials up-to-date is too often lacking
- No single entity has a mandate to organize these disparate efforts into a collective effort whose impact would be much greater than the sum

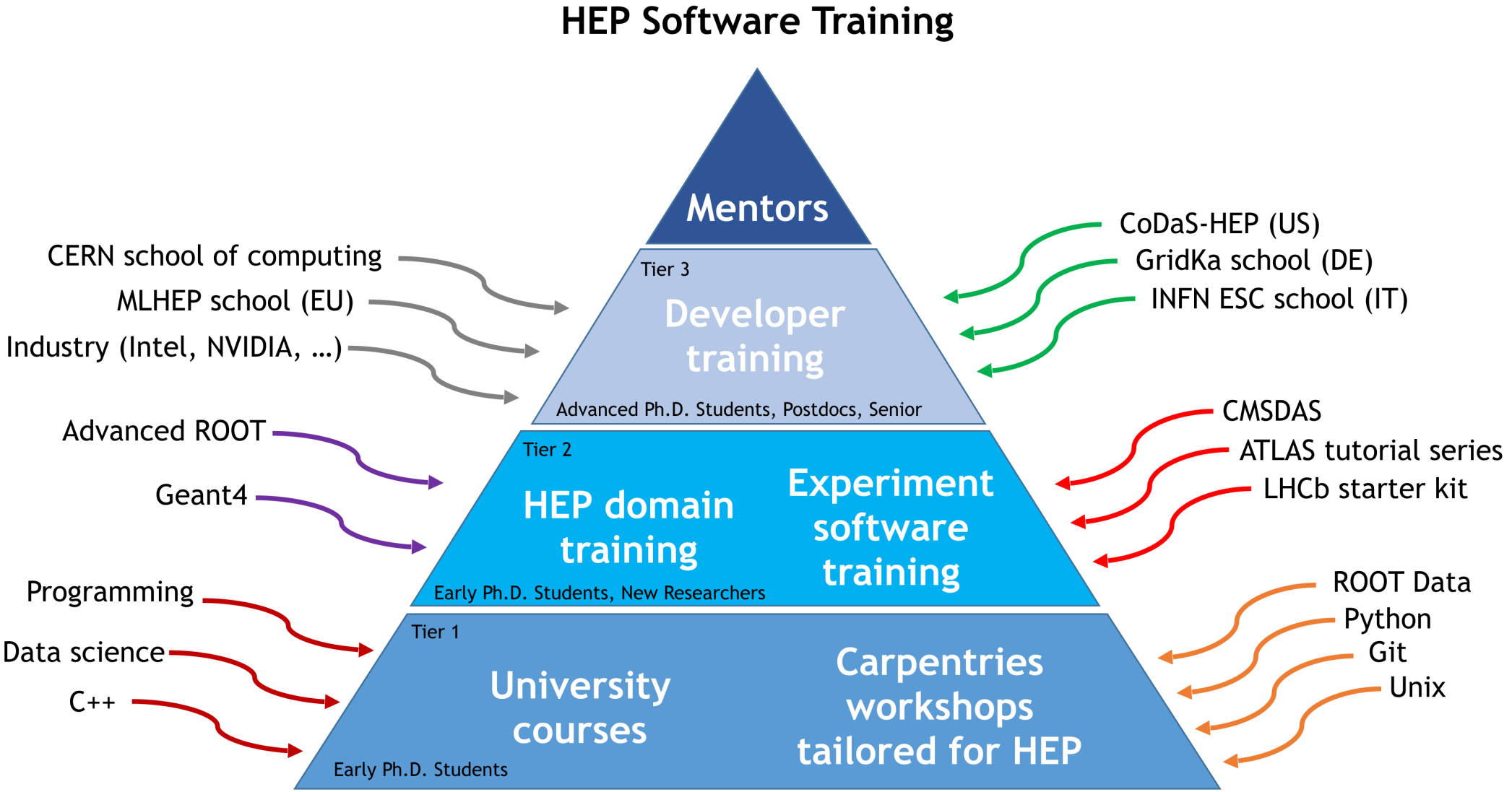
Examples of Current Training

- Experiment Specific
  - CMS - CMS Data Analysis School (CMSDAS) - help new collaborators learn hands-on the complex set of software tools needed to work with CMS data and the distributed CMS computing system
  - LHCb and ALICE - joint basic entry level software training called StarterKit, Belle II has its own flavor of StarterKit
  - ATLAS - tutorials setup for beginner, intermediate, developer level training via talks/recorded videos/workbook
  - Neutrinos, Muons, Astro,DES community at FNAL Currently have Online Software Exercises and 1-2 day workshops aligned with collaboration meetings
- Non-experiment
  - INFN School on Architectures, Tools and Methodologies for Developing Efficient Large Scale Scientific Computing Applications (ESC)
  - CoDaS-HEP: School on tools, techniques Data Science for HEP

Training Vision

- Establish a *community framework for software training* in order to prepare the scientific and engineering workforce required for the computing challenges of HEP experiments
- For sustainability , *build "users" community* as well as *"developer" community*
- Instead of new curriculum development *leverage and build upon existing material* from the HEP and larger research community
- *Build training activities and material into a "common good"* with a strong community of both instructors and participants, and with a feeling of *community ownership*

Training Chain



Upcoming activities

- Co-organize (with IRIS-HEP) very first introductory level Software Carpentry (git/python/PyRoot/uproot) workshop at Fermilab (1-2 April 2019)

- Co-organize CoDaS-HEP School (Princeton) - July 2019
  - tools, techniques and methods for *Co*mputational and *Da*ta *S*cience for High Energy Physics

- Introductory “first test run” level Carpentry session for US-ATLAS at LBNL and CMS/ATLAS at CERN, the latter connecting to the international parts of the experiment, build on the experience of the FNAL workshop.

- Invite LHCb/ALICE to showcase Starterkit on pre and first analysis steps

Links/Info

- IRIS-HEP website http://iris-hep.org/
  - Jobs on IRIS-HEP and Collaborating Projects http://iris-hep.org/jobs
  - General public announcement mailing list for IRIS-HEP events, talks, meetings, workshops, opportunities for training and job opportunities (subscribe to)
  - announcements@iris-hep.org

- HSF (HEP Software foundation) - https://hepsoftwarefoundation.org
  - General Information about HSF (subscribe to): hsf-forum@googlegroups.com
  - Discussions and activities in the HEP Software Foundation mailing lists can be found here (General and Dedicated Forums): https://hepsoftwarefoundation.org/forums.html
  - You can contribute https://hepsoftwarefoundation.org/cwp/cwp-working-groups.html
  - HSF Events/Workshops - https://hepsoftwarefoundation.org/events.html