PyHEP Workshop Summary and Outcomes

Eduardo Rodrigues and Graeme Stewart

2018-07-08

Overview

- Python is on an upward trajectory
  - Data science, Machine learning providing strong drivers
- HEP usage is increasing too
  - Coupled to expansion of Python ecosystem, but also building on Python's traditional strengths
- Notebooks are a huge hit
  - Many thanks to Vidar for the JupyterLab talk
- Pre-workshop questionnaire
  - Training
  - Plotting
    - Galleries are really useful to find examples, didactic too!
  - Installation

Inventory

- Even we did not know what useful packages are available
- Inventory of tools appropriate to HEP would be great
  - With notebooks and galleries to show how to use them
- Orphaned packages, but still useful?
  - Way to look for a maintainer
  - Scikit-HEP has handed over packages between maintainers
- Repository of expertise in the PyHEP community
  - Ties well with hot topic of education and training across the field

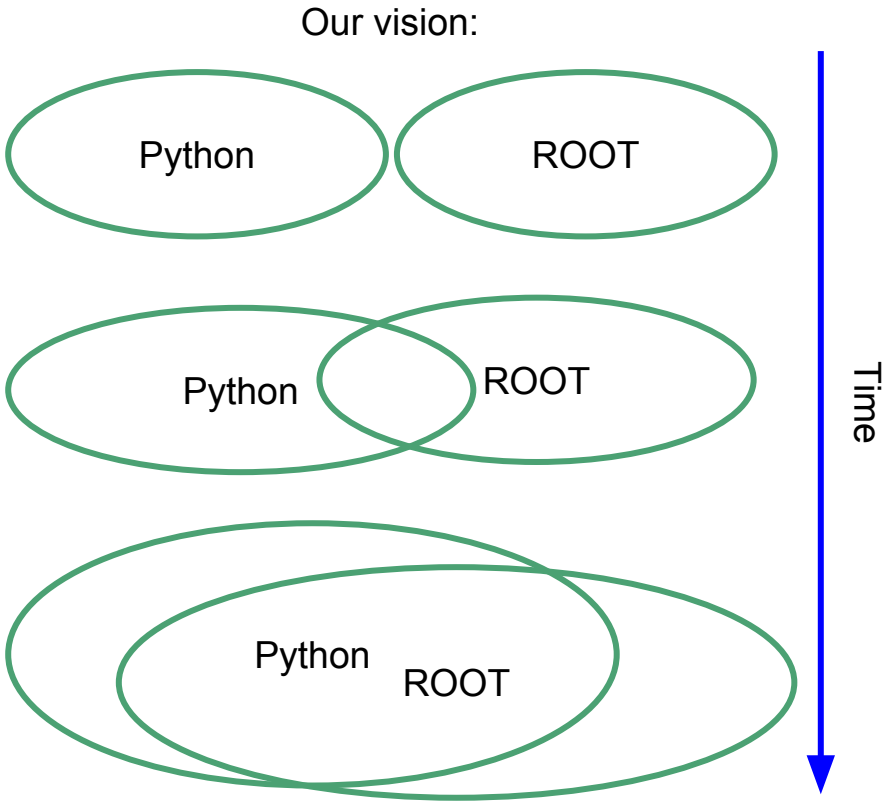
Experiment and Analysis Directions

- Python is a language that can be used for all the computing
  - End to end
  - Flexible
  - Naturally modular
- The Belle II analysis/training jupyter cluster looks great
- Extensions to full analysis clusters?
  - Need good integration with storage
  - SWAN as an integrated, stable and reproducible environment

This looks like a key direction, aligned with Community White Paper Roadmap

ROOT

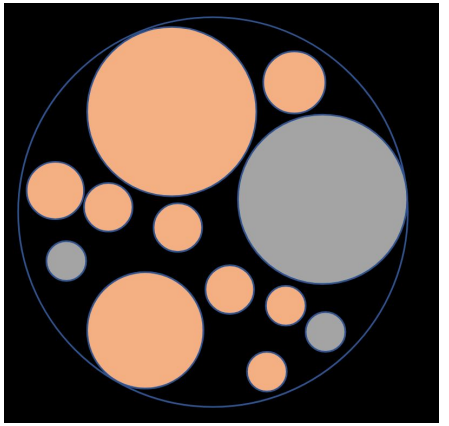
- Data model ideal for HEP
- Fitting, histograms best in class
- Heavy component
  - Too burdensome for some small experiments it seems
  - Modularity would help
- Easier ways to install
  - NLeSC effort was greatly appreciated
- cppyy is a contribution that is far less well known than it should be
- PyROOT developments exciting
  - Particularly adding pythonisation, to make things natural



Training

- The Belle II analysis jupyter cluster looks great
- Extensions to analysis clusters?
  - Need good integration with storage
  - Cf. SWAN
- Using standard Python libraries to achieve HEP workflows is a concern for *our community*
- Training session discussion was brilliant, right!
  - (What do you mean you forgot already what was said?)

Distribute and Install



- Can we be as standard as possible?
  - CMS using pip + PyPI
- Distributing whole HEP stack is a difficult problem
  - Worse than Python, multidimensional
- Distinguish experiment stack from analysis
  - Toolboxes, not frameworks
  - SWAN encapsulates things really well
  - Daring view: ubiquitous network access + browser...
- Modularity and flexibility of the solution vital
  - HSF Packaging Group should pay more attention to this

To Python 3

- Will be painful for the large pieces, but we just have to do this
  - LS2 project for LHC experiments - having to get to the end of Run 3 with an unsupported Python would be uncomfortable
  - An increasing gap between legacy Python 2 and Python 3 would hurt

Next Steps

- More workshops like this?
  - Format to be defined - format itself, duration, location depending on format
  - Adding...
    - Training?
    - Hackathons?
- Community inventory/information
  - Meshes with HSF Software Project inventory
- HSF Forum list is great for general announcements; HSF Tech Forum for technical topics
  - These are your lists, please use them!
- Do post on the HSF coordination mailing list for a smaller discussion
  - Or get in touch privately if you prefer
- Do we want a PyHEP list?

Thank You!

This was a great workshop, we look forward to more in the future