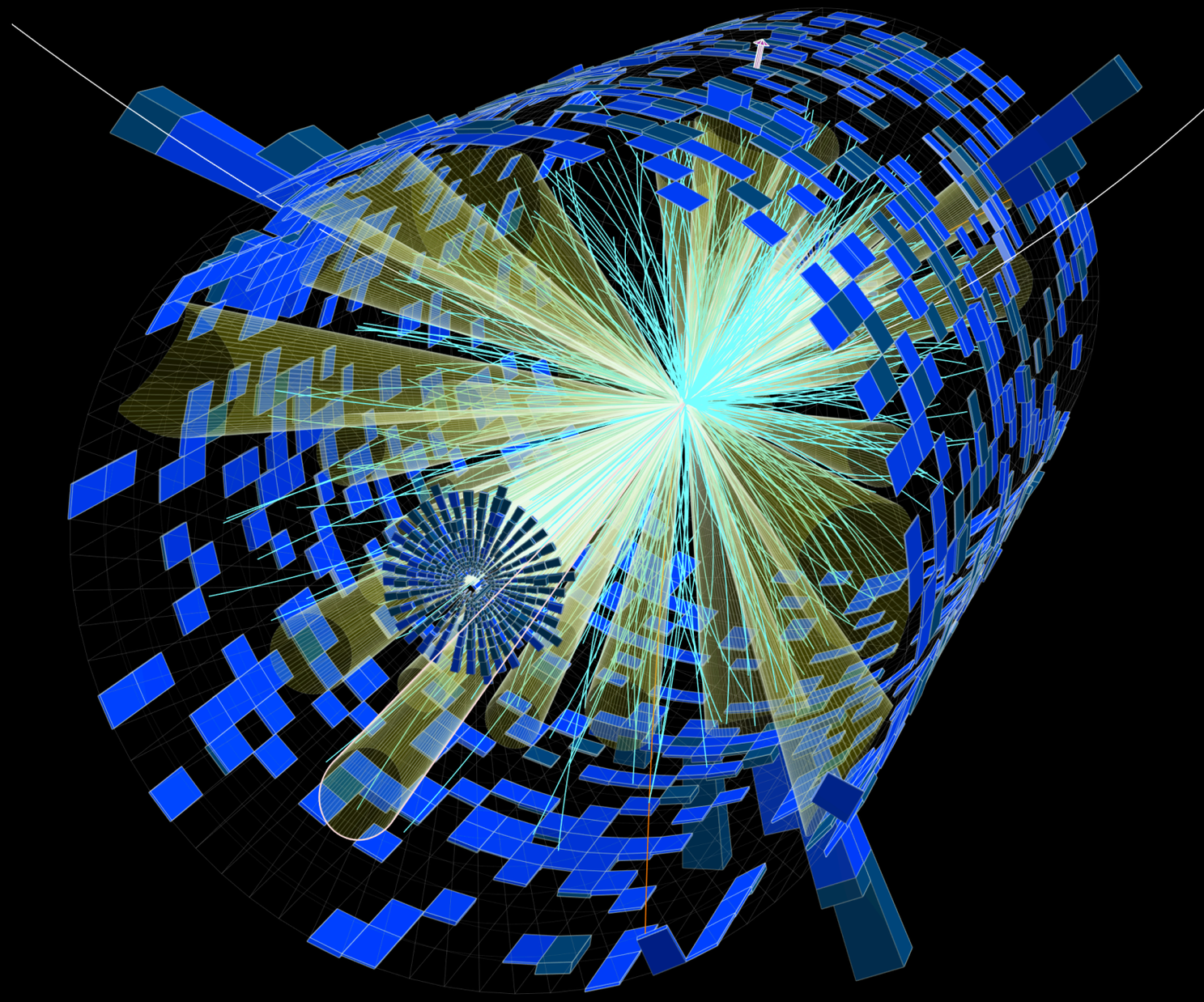
NYU Center for Data Science

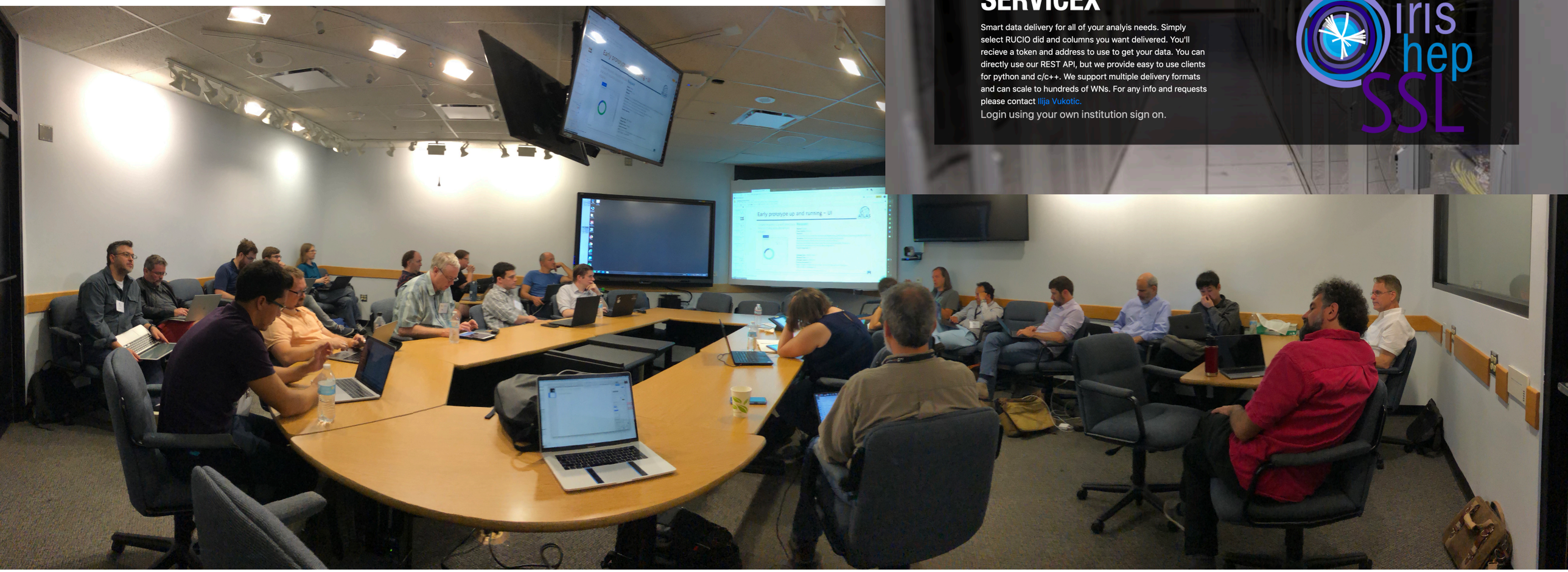
Center for Cosmology and particle physics

ANALYSIS SYSTEMS CLOSEOUT



@KyleCranmer
New York University
Department of Physics
Center for Data Science
CILVR Lab

ServiceX Demo



Marc Weinberg demo of ServiceX with Kafka, Kabana, awkward, pyroot

- Transformer reads ROOT produces awkward array.
  - pyroot transformer slow, breakout about using func adl C++ code generator to make transformer fast
- Analyzer jobs ingest awkward array. job, has been Coffea+spark
- Components Distributed across CERN, GKE, and River

Biggest challenge: FermLab's firewall. Outcome: Z boson still exists!

ROADMAP FOR STATS TOOLS

Key

Specifications

Concrete Implementations

Possible Intermediate Representation

Common API

Downstream Stats tools (like RooStats)

HistFactory

Combine card

Decay Language

StatisFactory

pyhf

RooFit

GooFit

zfit

New idea [long term] Intermediate Representations, Conversions, Optimizations

pyfitcore API

(name under discussion)

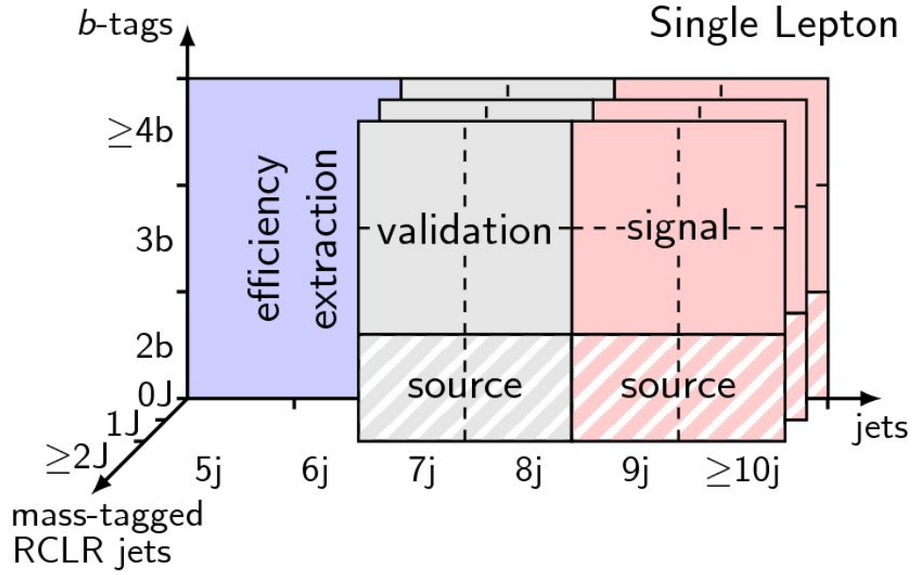
Downstream Stats tools (like RooStats)

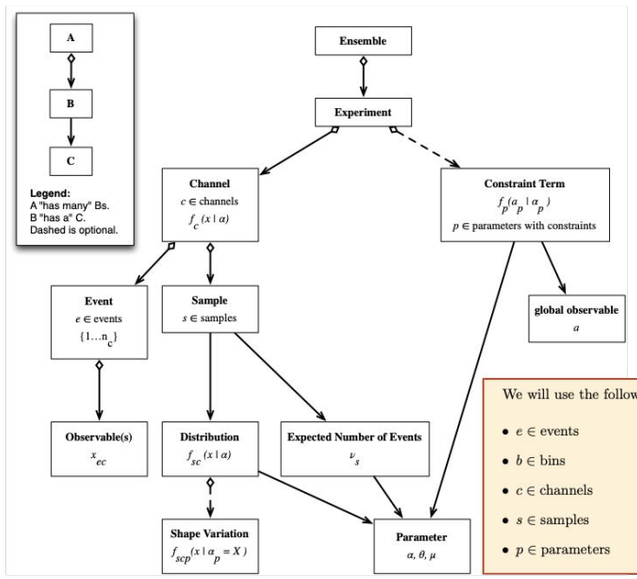
BORDER INDICATES NEW PROJECTS

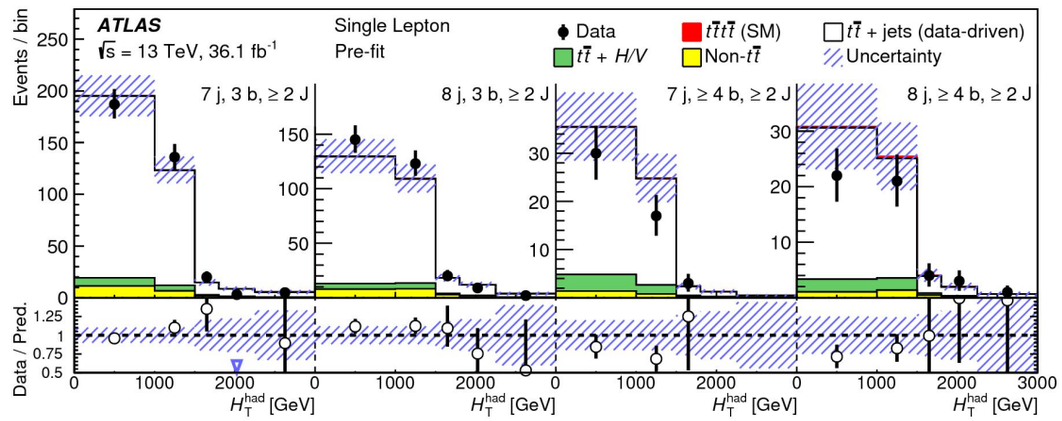
A point of convergence

Several aspects of Analysis Systems converge in a typical physics plot:

- Specification of signal / validation / control regions
- Specification of variables to be used for stat analysis
- Reduction to that format running on data and MC
- Management of MC samples, data driven backgrounds, etc.
- Management of systematic variations
- Feed reduced data (eg. histograms) into specification for statistical model / likelihood function
- Fitting & statistical tools
- Publishing results & derived data products
- Analysis preservation & gateways targeting reinterpretation







FUTURE PLANS

Integrate Coffea -> pyhf a la HistFitter / TRexFitter

- Use the event / column languages to define
  - Variable definitions used to define the observable and variables used for event selection
  - Predicate used for event selection that defines the “channel”
- Alex Held will be working on this
- Some initial prototyping: https://github.com/lukasheinrich/pyhfinput

Going beyond HistFitter and TRexFitter

- Inferno-style
  - replace static observable with differentiable computational graph to compute it
- MadMiner / GPs etc.
  - Replace histogram representation of distribution with ML-based density estimation

...AND MUCH MORE

boost-histogram and hist Roadmap

Henry Schreiner

September 12-13, 2019

iris hep
Institute for Research & Innovation in Software for High Energy Physics

PRINCETON UNIVERSITY

gordonwatts / func_adl_electrons

Watch 1 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Security | Insights

Branch: master | func_adl_electrons / ElectronData.ipynb | Find file | Copy path

gordonwatts First version at getting the data a33817c 6 hours ago

1 contributor

204 lines (203 sloc) | 19.4 KB | Raw | Blame | History

Fetching Simple Electron Data

This notebook is a demo for using the various things to feetch electron data. We are copying a request from the `ServiceX` system. Our backer

Here is the original `ServiceX` request:

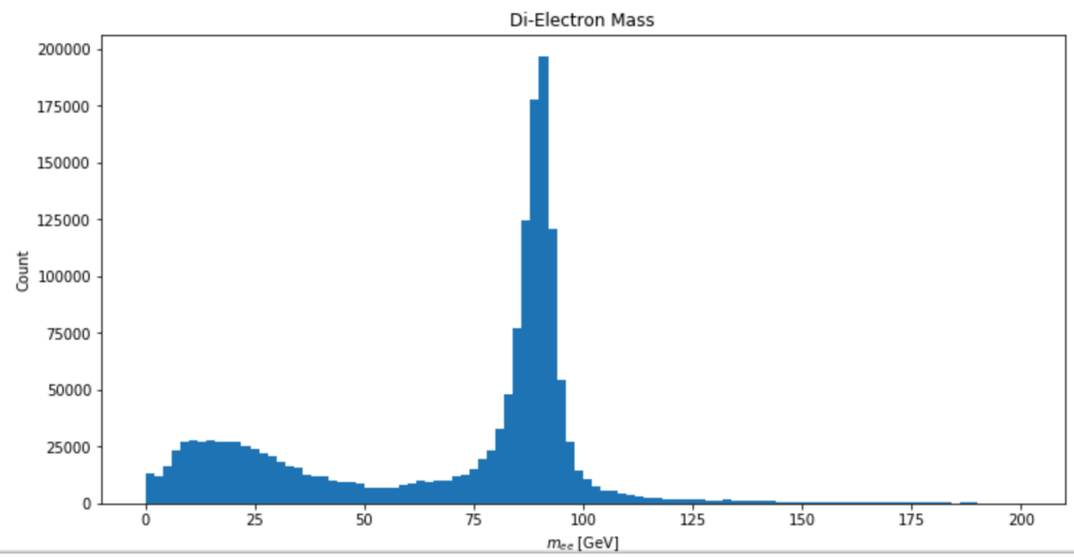
```
In [8]: leptons_per_event = await leptons_per_event_as

Files that were returned:
  ['file:///C:\\Users\\gordo\\Documents\\func-adl-cache/40ae9bb8bd6cf8bcb7ae703c715939d7/ANAL
001.root', 'pandas_tree28']

In [15]: v_particles = uproot_methods.TLorentzVectorArray.from_ptetaphi(
             leptons_per_event[b'ElePt'], leptons_per_event[b'EleEta'],
             leptons_per_event[b'ElePhi'], leptons_per_event[b'EleE'],
             )

In [17]: v_particles = v_particles[v_particles.counts >= 2]
         diparticles = v_particles[:, 0] + v_particles[:, 1]

In [39]: plt.figure(figsize=(12, 6))
         plt.hist(diparticles.mass/1000.0, bins=100, range=(0,200))
         plt.title('Di-Electron Mass')
         plt.xlabel('$m_{ee}$ [GeV]')
         plt.ylabel('Count')
         plt.show()
```



Reviewed and discussed Y2 milestones and planning

- Generally in pretty good shape
- Lots to do!

No more time to create slides