

DeepJet Framework

Swapneel Mehta, Mauro Verzetti, Jan Kieseler, Markus Stoye, Maurizio Pierini
CMS Experiment, EP-CMG-PS
CERN

Large Hadron Collider

CMS
experiment

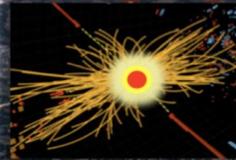
Large Hadron Collider

- 27 km (17 miles) circumference
- 1600 superconducting magnets at 1.9° K (-271.3° C or -459.7° F)
- 120 tonnes of liquid helium
- Accelerates beams of protons to 99.9999991% the speed of light



Large Hadron Collider

CMS



Large Hadron Collider

- Proton beams circulate 11,245 times/sec
- 100's of millions of proton-proton collisions/second
- Collisions are a billion times hotter than the centre of the sun and create new particles ($E = mc^2$)

CMS Centre @ CERN

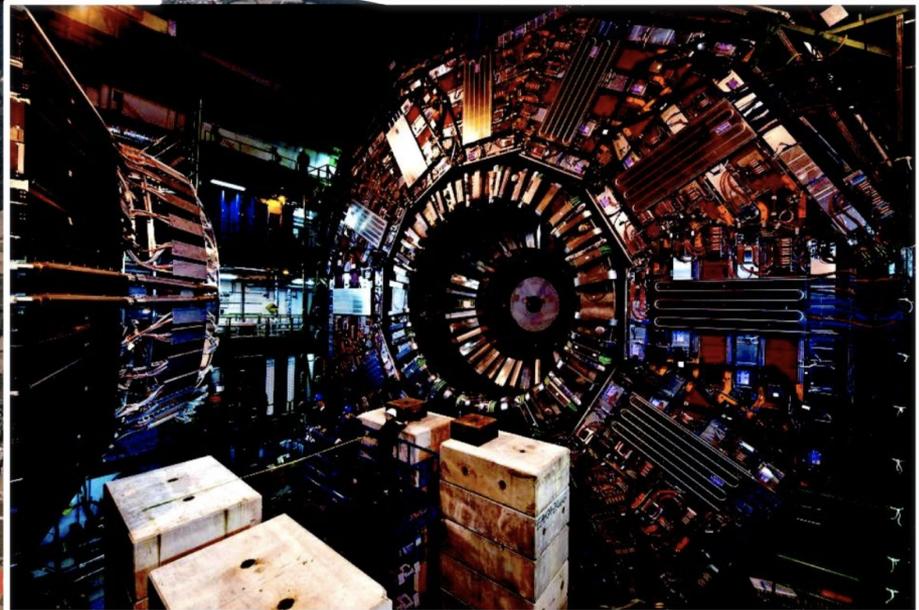
CERN Computer Centre

Compact Muon Solenoid (CMS)

CMS

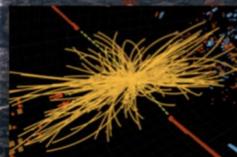
Large Hadron Collider

- Detects new particles created in LHC collisions
- 21m long and 15m high in a huge cavern 100m underground
- 12,500 tonnes



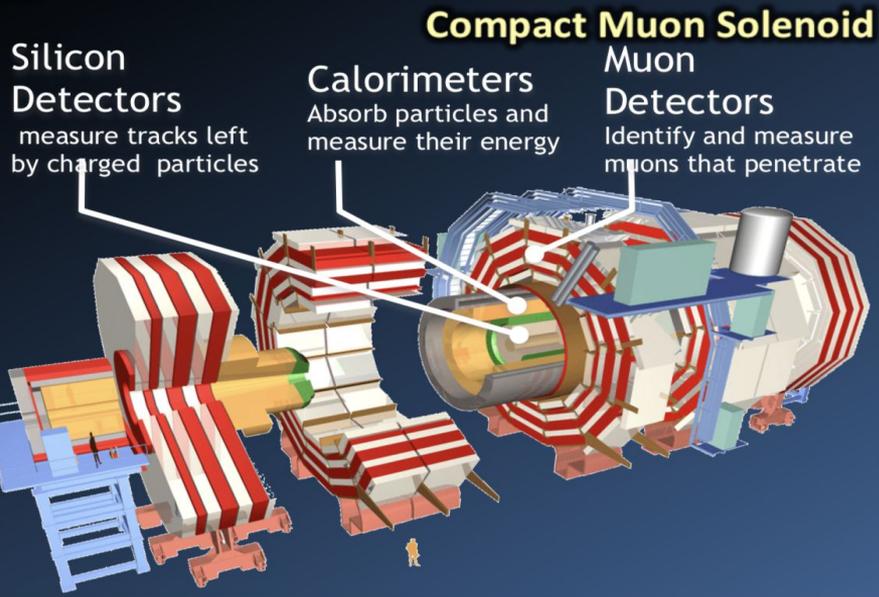
Compact Muon Solenoid (CMS)

CMS

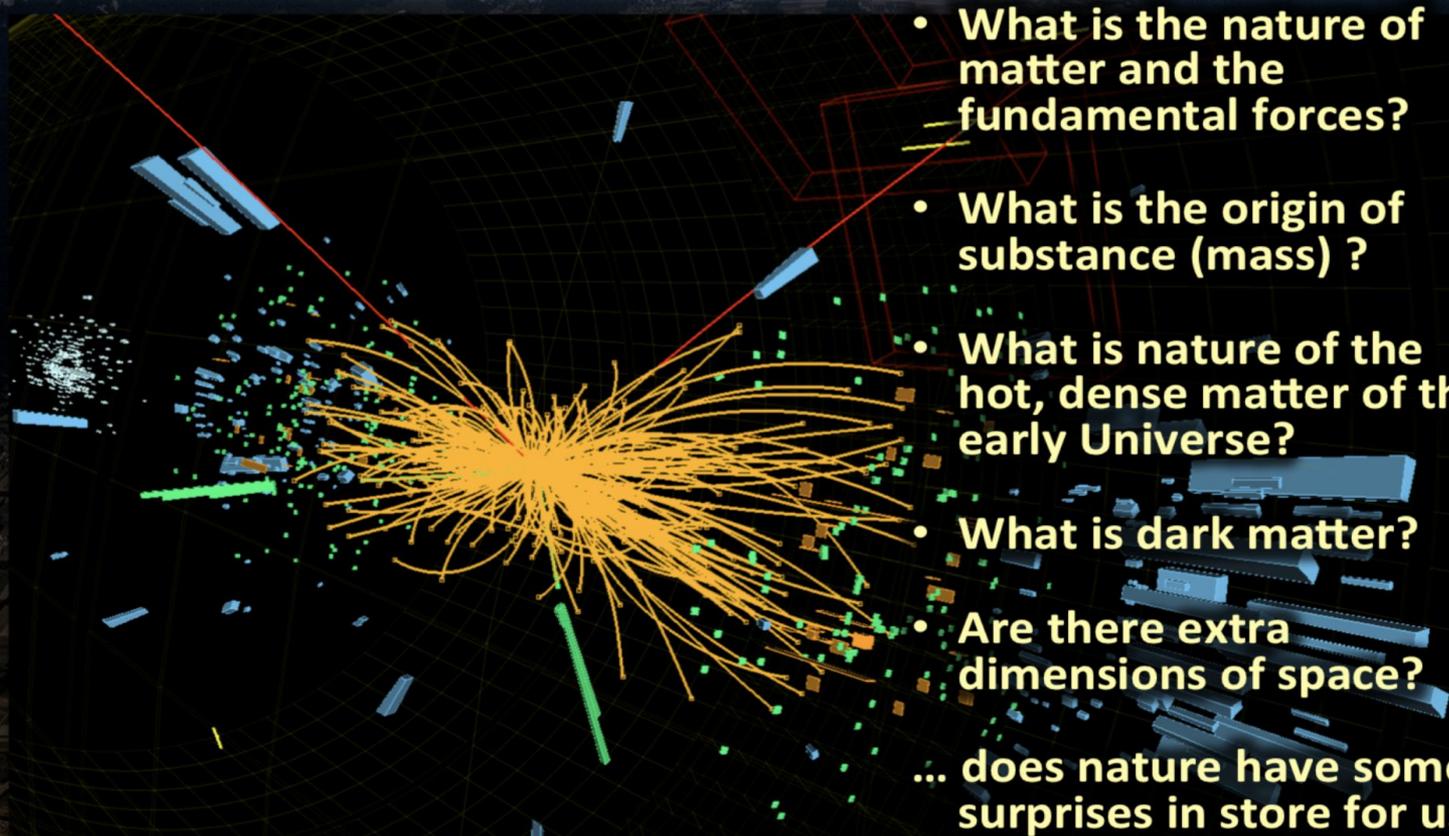


Large Hadron Collider

- Measures particle types, energies and directions
- Selects and records a few 100 interesting events/sec



CMS physicists study LHC collision events to learn about the Universe in which we live



- What is the nature of matter and the fundamental forces?
- What is the origin of substance (mass) ?
- What is nature of the hot, dense matter of the early Universe?
- What is dark matter?
- Are there extra dimensions of space?
- ... does nature have some surprises in store for us?

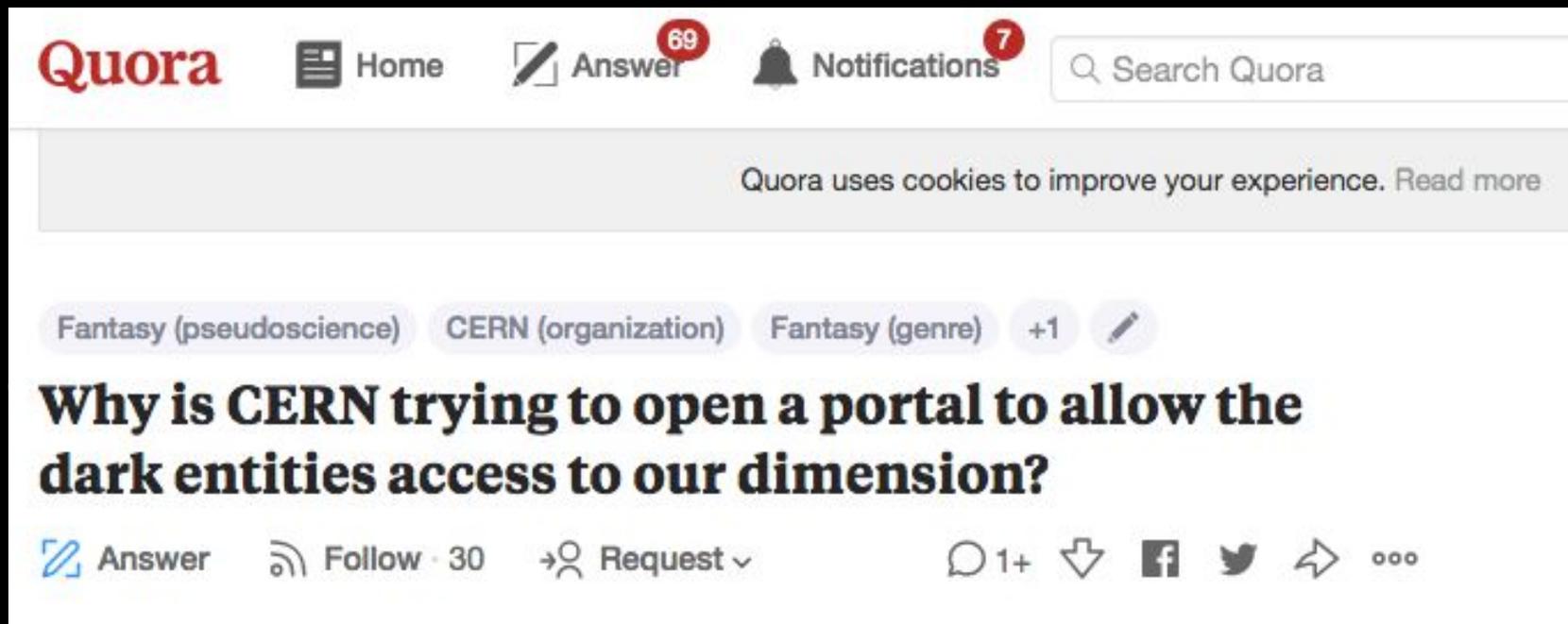
Machine Learning

1. Comprehensive libraries
2. Fantastic documentation
3. Interactive Tutorials
4. Developer Community Support



Why build a
library designed
for high-energy
physics?

Computer Scientists don't always understand requirements for particle physics...



The image shows a screenshot of a Quora question page. At the top, the Quora logo is on the left, followed by navigation links for Home, Answer (with a red badge showing 69), and Notifications (with a red badge showing 7). A search bar is on the right. Below the navigation is a grey banner stating "Quora uses cookies to improve your experience. Read more". Underneath the banner are several topic tags: "Fantasy (pseudoscience)", "CERN (organization)", "Fantasy (genre)", and "+1" with a pencil icon. The main question is "Why is CERN trying to open a portal to allow the dark entities access to our dimension?". Below the question are interaction options: "Answer", "Follow · 30", "Request" (with a dropdown arrow), and social sharing icons for comments (1+), a download icon, Facebook, Twitter, a share icon, and a menu icon (three dots).

Quora

Home Answer ⁶⁹ Notifications ⁷ Search Quora

Quora uses cookies to improve your experience. Read more

Fantasy (pseudoscience) CERN (organization) Fantasy (genre) +1

Why is CERN trying to open a portal to allow the dark entities access to our dimension?

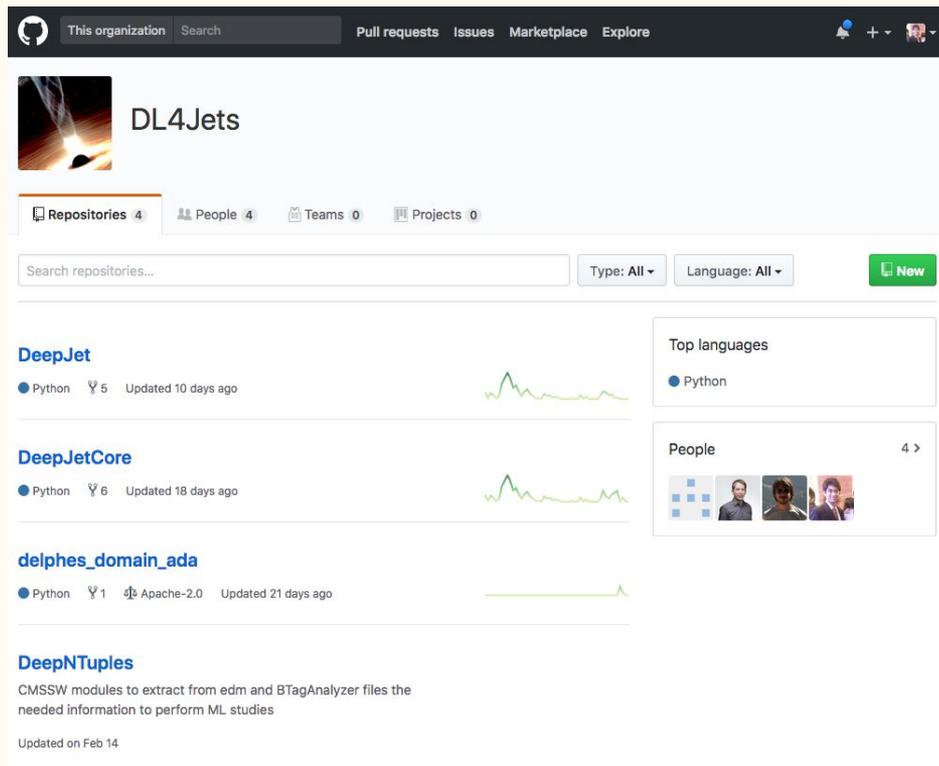
Answer Follow · 30 Request ▾ 1+ [Facebook] [Twitter] [Share] [More]

Physicists don't always write great code...

```
public Date getNextDay() {  
    try {  
        Thread.sleep(TimeUnit.DAYS.toMillis( duration: 1));  
        return new Date(); //success  
    } catch (InterruptedException e) {  
        e.printStackTrace();  
        return null; //failure  
    }  
}
```

Best of Both Worlds

1. Implement fast, efficient machine learning algorithms for physics
2. Provide high-level functions/wrappers for low-level tasks
3. Handle common bottlenecks - esp. memory -related issues
4. Create an extensible, easy-to-use framework

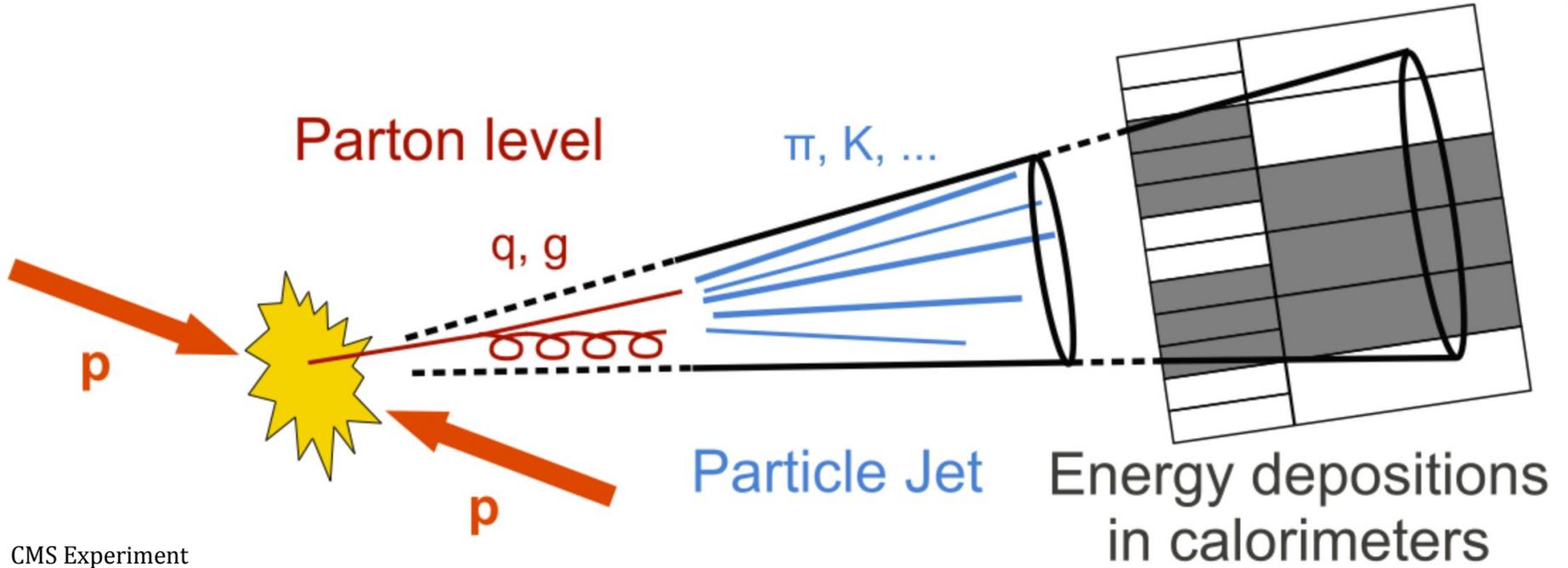


The screenshot shows the GitHub organization page for DL4Jets. The header includes navigation links for Pull requests, Issues, Marketplace, and Explore. The organization's profile picture is a stylized image of a particle detector. Below the profile, there are statistics for Repositories (4), People (4), Teams (0), and Projects (0). A search bar for repositories is present, along with filters for Type and Language, and a 'New' button. The main content area lists several repositories:

- DeepJet**: Python, 5 forks, updated 10 days ago. Includes a green line graph showing activity.
- DeepJetCore**: Python, 6 forks, updated 18 days ago. Includes a green line graph showing activity.
- delphes_domain_ada**: Python, 1 fork, Apache-2.0 license, updated 21 days ago. Includes a green line graph showing activity.
- DeepNTuples**: CMSSW modules to extract from edm and BTagAnalyzer files the needed information to perform ML studies. Updated on Feb 14.

On the right side, there are sections for 'Top languages' (Python) and 'People' (4 members).

So what exactly is
Jet Physics?

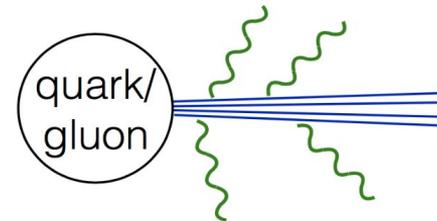
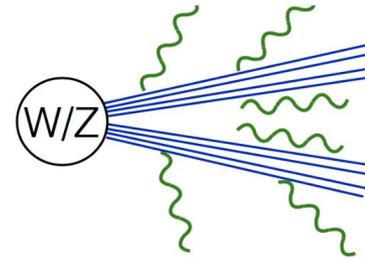


CMS Experiment

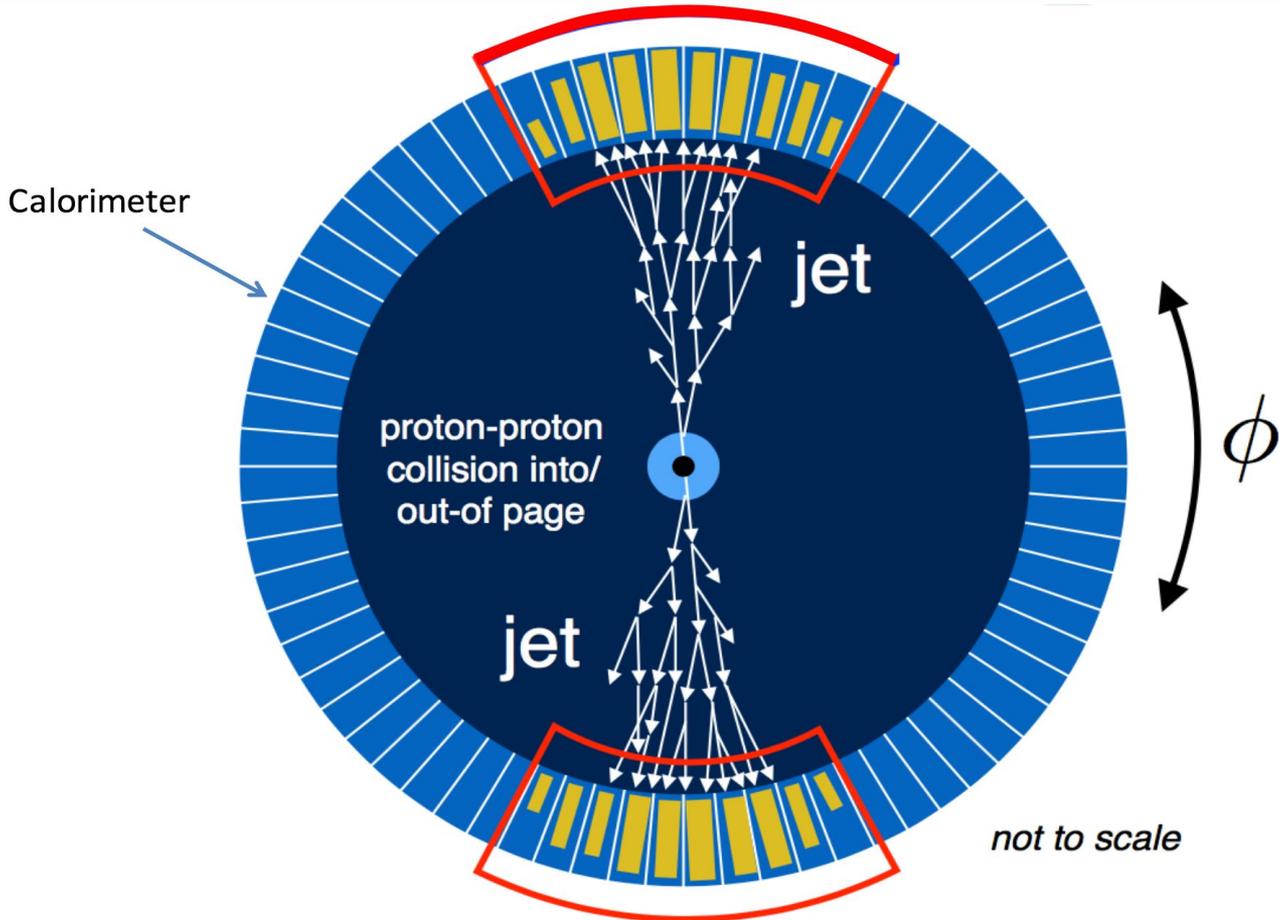
Jets: Collimated Streams of Particles

- How to deal with the high dimensionality and complexity of Jets?
- Example, suppose we want to classify jets

$$\frac{P(\text{jet}|\text{class} = W)}{P(\text{jet}|\text{class} = \text{Quark})}$$



- How can we learn probability distributions (or ratios) for jets?



What does this
library do?

Features of DeepJet

- Data Conversion
 - Model Training
 - Prediction
 - Model Evaluation
-

- File-by-File
- Avoids memory threshold crossed (EOS)
- Handles user-defined data structures
- Preprocessing support
- Parallelized operation

Conversion

- Keras-wrapped Tensorflow backend
- Additional callbacks
- Monitor validity of tokens
- Bookkeeping support

Training

- Create compatible prediction data structures
- Support for Plots
- Export of models and data structures

Prediction and Evaluation

Yeah, but why
should I use it?

- Modularised code, easy to understand
- Templates for quick-start
- Step-by-step documentation
- Elaborate examples and use-cases

Simplicity

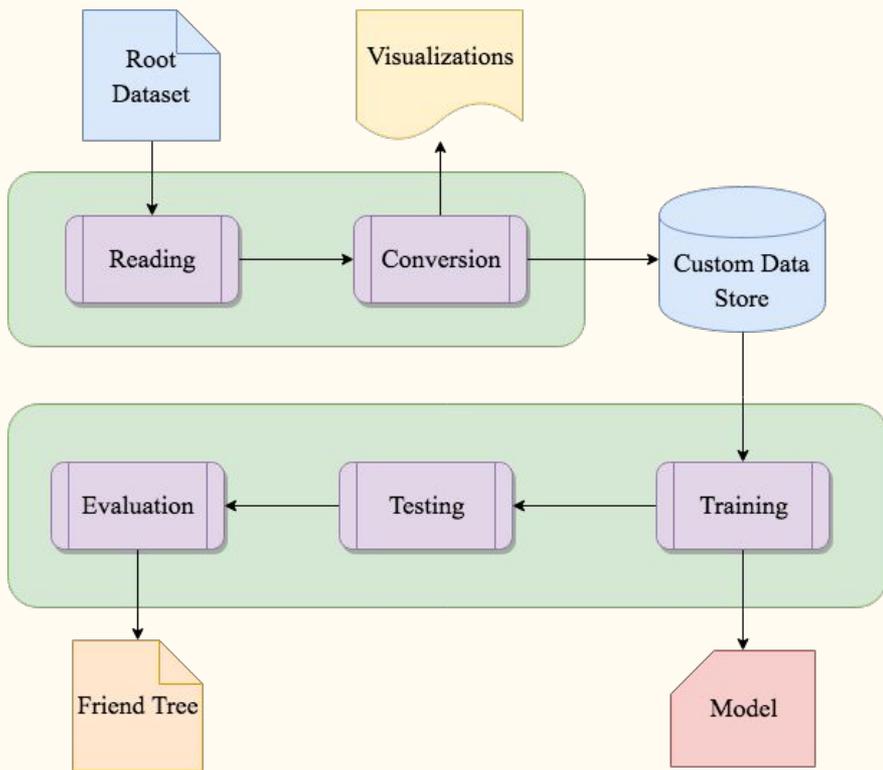
- Custom CPP Extensions
improve efficiency for
Python
- Automation of specific
tasks
- Anaconda Environment

Support

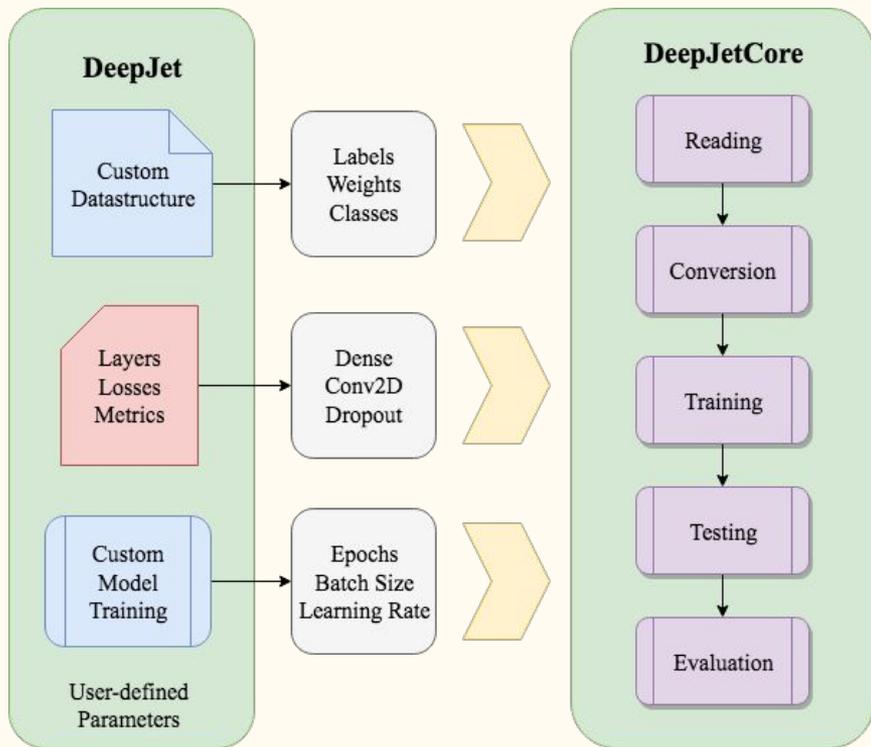
- Available as a pip package for Python 3.6
- Tensorflow 1.8 supported
- Integrating support for TFRecords
- Docker Image Distribution

Upgrades

Interesting! Tell
me more about
this library



DeepJetCore



DeepJet

DeepJet Demo



Conclusion

- Easy-to-use Framework
 - Faster conversion and training
 - Diverse use-cases
 - Scalable to large datasets
-

Want to learn more
about Machine
Learning for
High-energy
Physics (MLHEP)?

Resources for Getting Started with MLHEP

<https://github.com/iml-wg/HEP-ML-Resources>

<https://www.coursera.org/learn/particle-physics>

[Shameless Plug]

<https://github.com/SwapneelM/awesome-particle-physics-for-non-physicists>

References

[Links]

- Lucas Taylor's CMS Experiment Slides
 - CMS Collaboration Public Outreach Slides
 - Dave Barney, Andre David CMS e-Masterclass Slides
 - Michael Kagan's Jet Classification Slides
-