DeepJet Framework

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

Large Hadron Collider

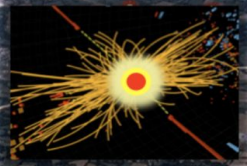
CMS experiment

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

Lucas Taylor, 2012

Large Hadron Collider



Large Hadron Collider

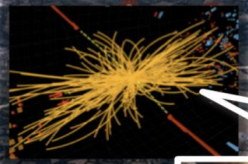
- Proton beams circulate 11,245 times/sec
- 100's of millions of proton-proton collisions/second
- Collisions are a billion times hotter than the centre of the sun and create new particles ($E = mc^2$)

CERN Computer Centre

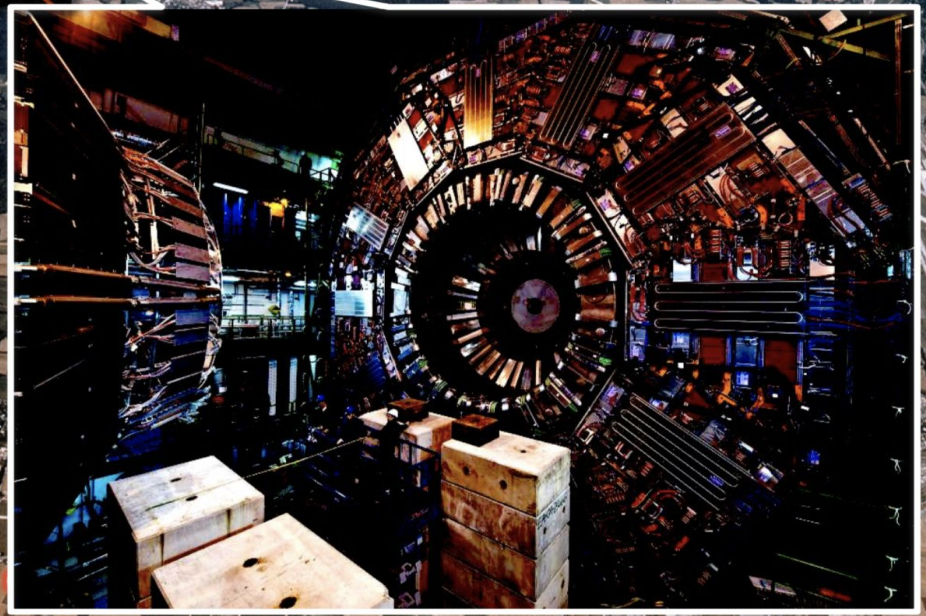
CMS Centre @ CERN

Compact Muon Solenoid (CMS)

CMS



Large Hadron Collider

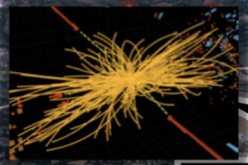


- Detects new particles created in LHC collisions
- 21m long and 15m high in a huge cavern 100m underground
- 12,500 tonnes

Lucas Taylor, 2012

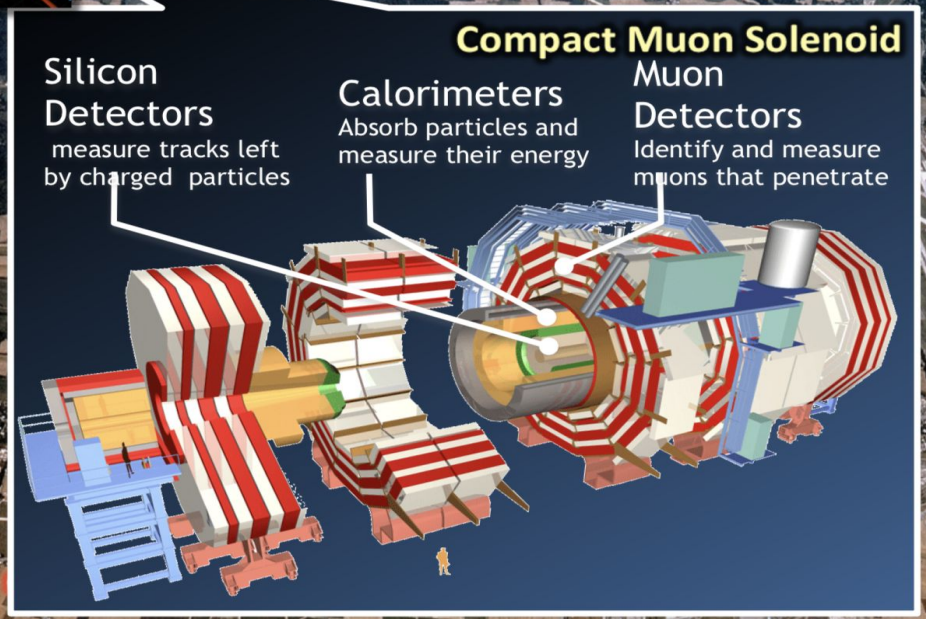
Compact Muon Solenoid (CMS)

CMS



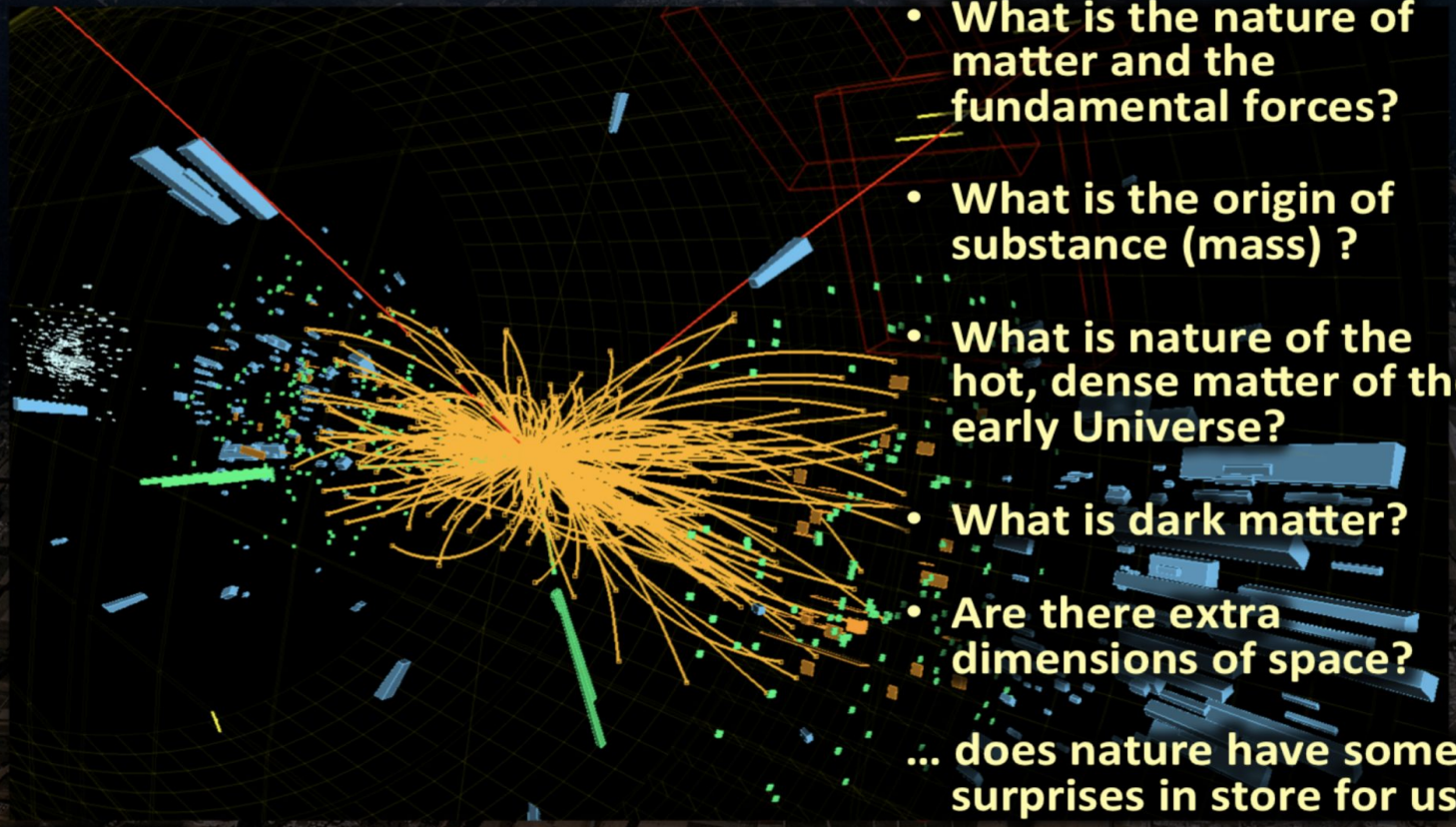
Large Hadron Collider

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



Lucas Taylor, 2012

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?

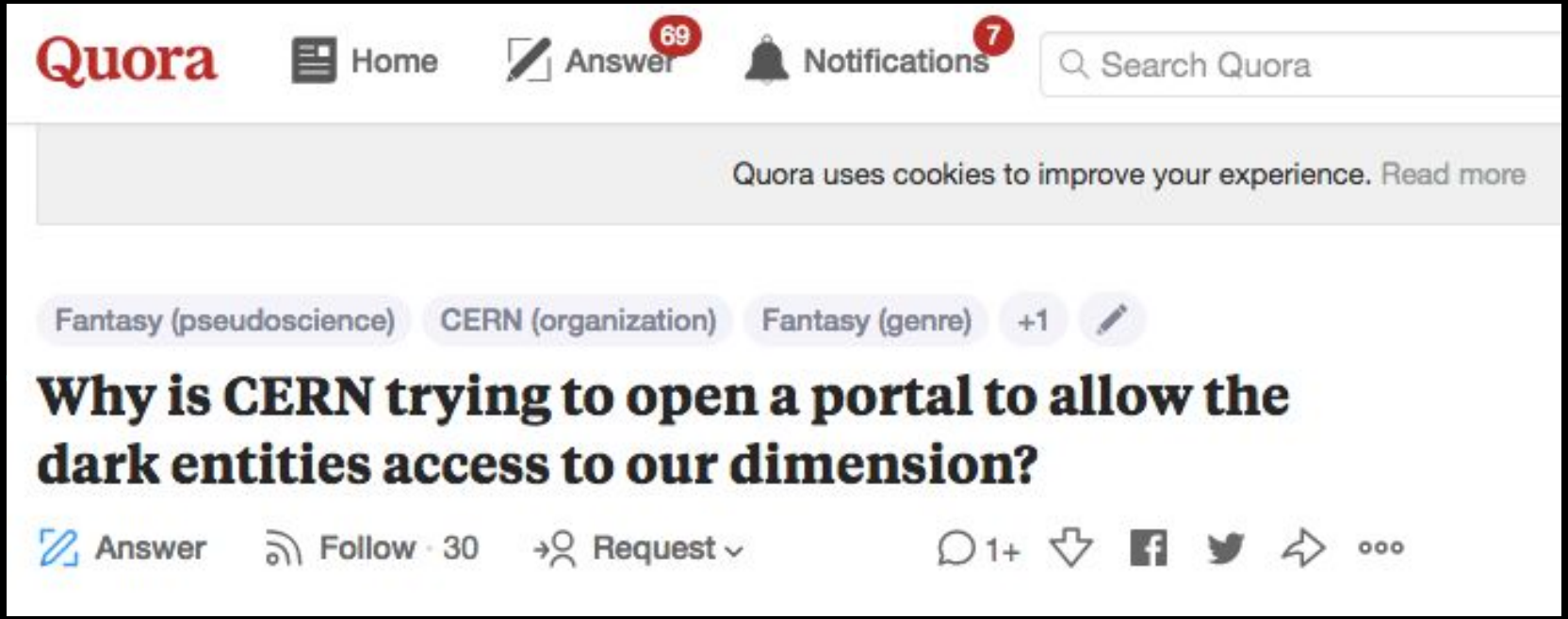
… does nature have some surprises in store for us?

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...

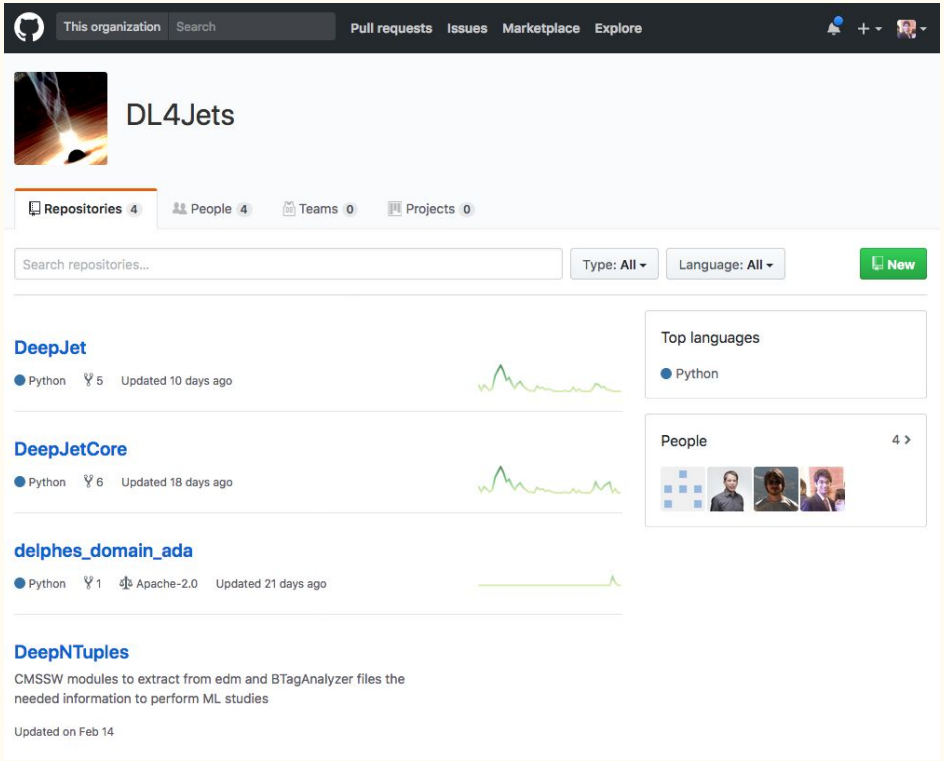


Physicists don't always write great code...

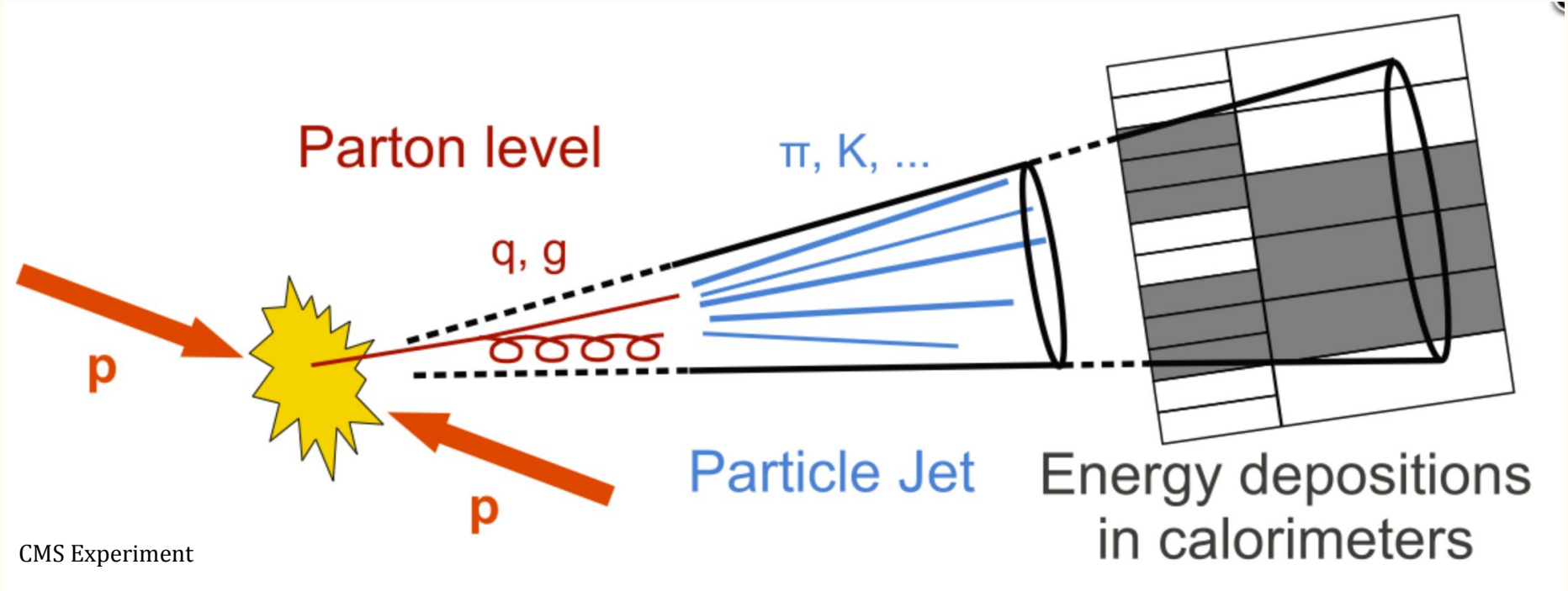
```
public Date getNextDay() {
    try {
        Thread.sleep(TimeUnit.DAYS.toMillis( duration: 1));
        return new Date();  //success
    } catch (InterruptedException e) {
        e.printStackTrace();
        return null;        //failure
    }
}
```

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



So what exactly is Jet Physics?



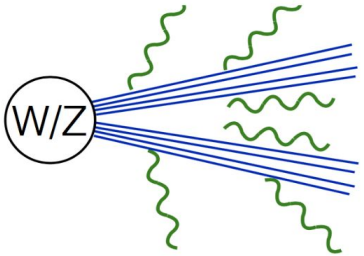
Jets: Collimated Streams of Particles

Jet Identification

- How to deal with the high dimensionality and complexity of Jets?

- Example, suppose we want to classify jets

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



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

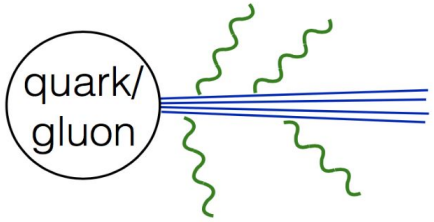


Image from A. Larkoski

Jets at the LHC

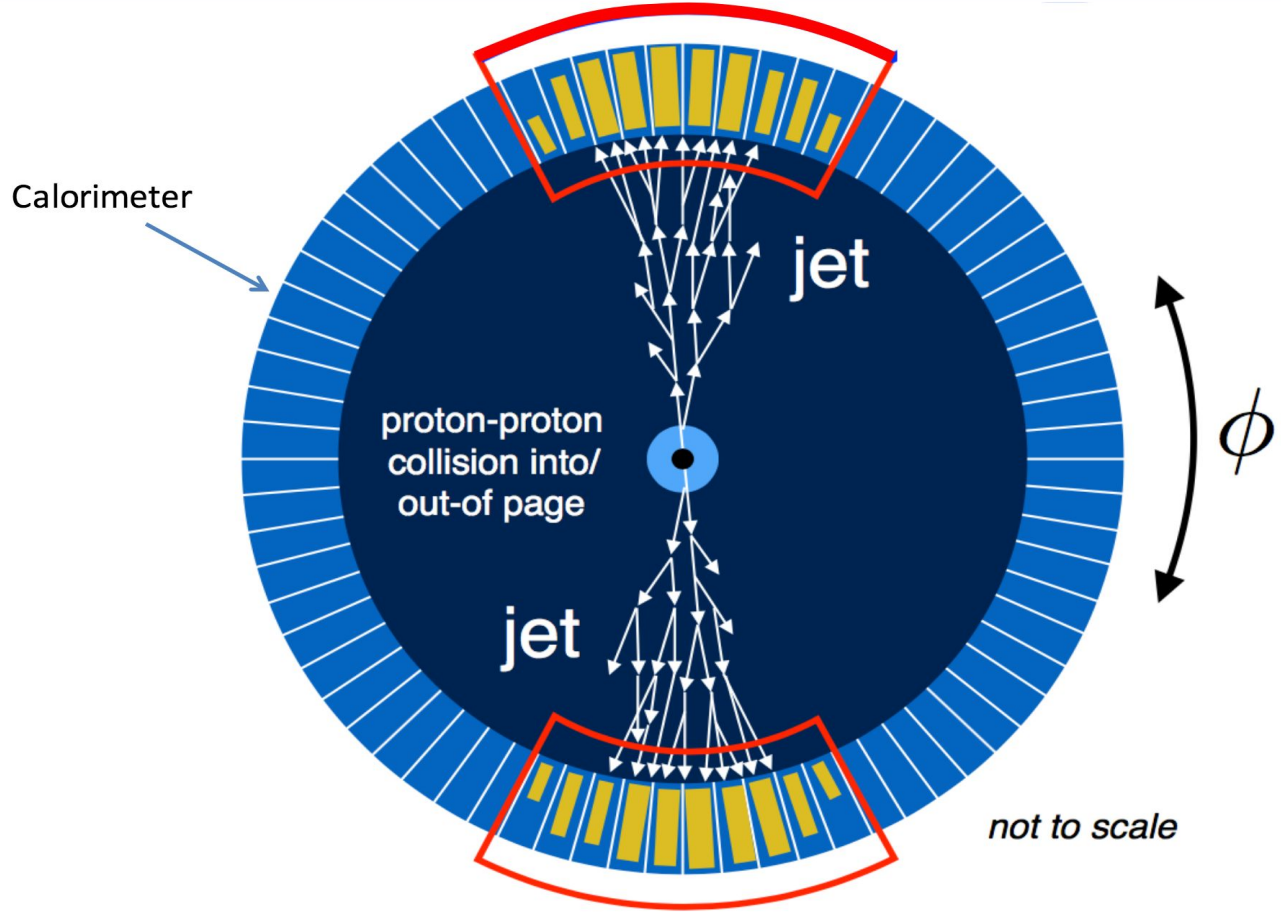


Image from B. Nachman

What does this library do?

Features of DeepJet

- Data Conversion
- Model Training
- Prediction
- Model Evaluation

Conversion

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

Training

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

Prediction and Evaluation

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

Yeah, but why should I use it?

Simplicity

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

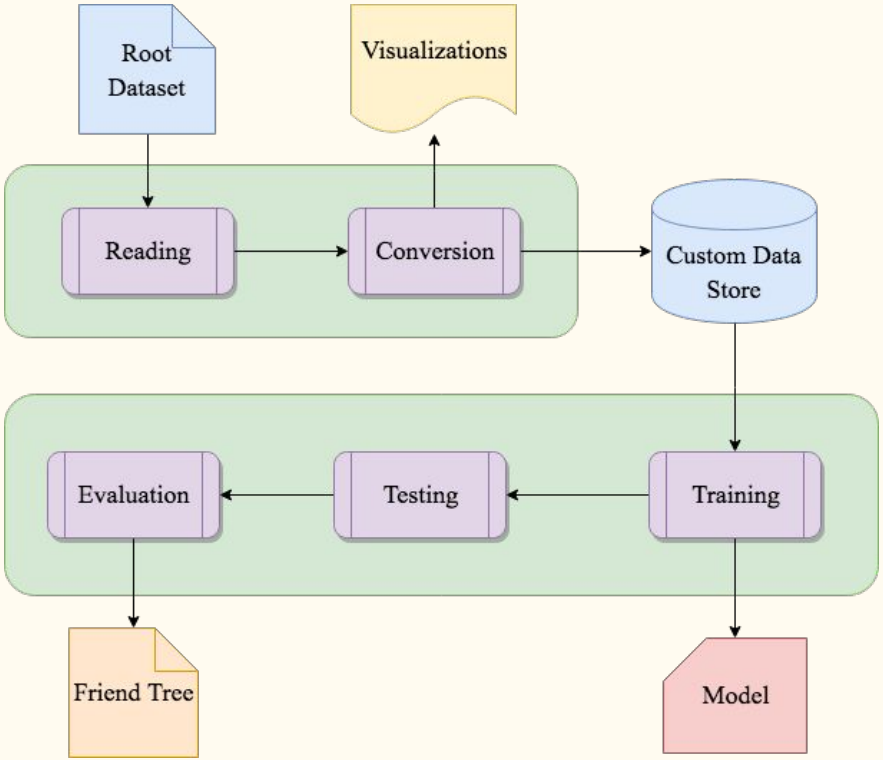
Support

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

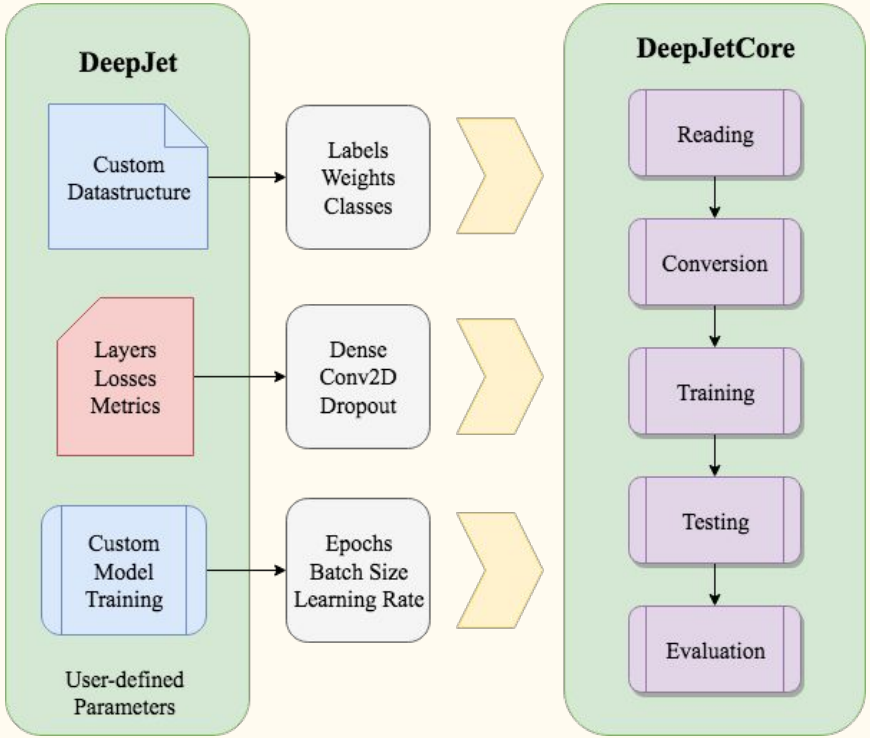
Upgrades

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

Interesting! Tell me more about this library



DeepJetCore


DeepJet
Custom Datastructure
Layers Losses Metrics
Custom Model Training
User-defined Parameters
Labels Weights Classes
Dense Conv2D Dropout
Epochs Batch Size Learning Rate
DeepJetCore
Reading
Conversion
Training
Testing
Evaluation

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