Yandex

Event Filter Update

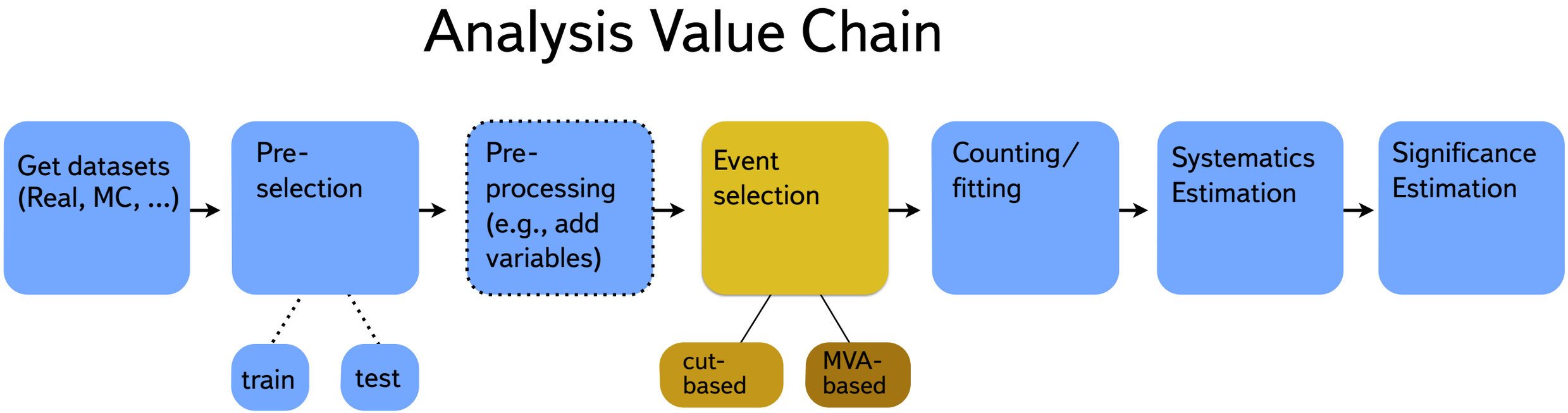
Andrey Ustyuzhanin

May 21 2014

What was Event Filter?

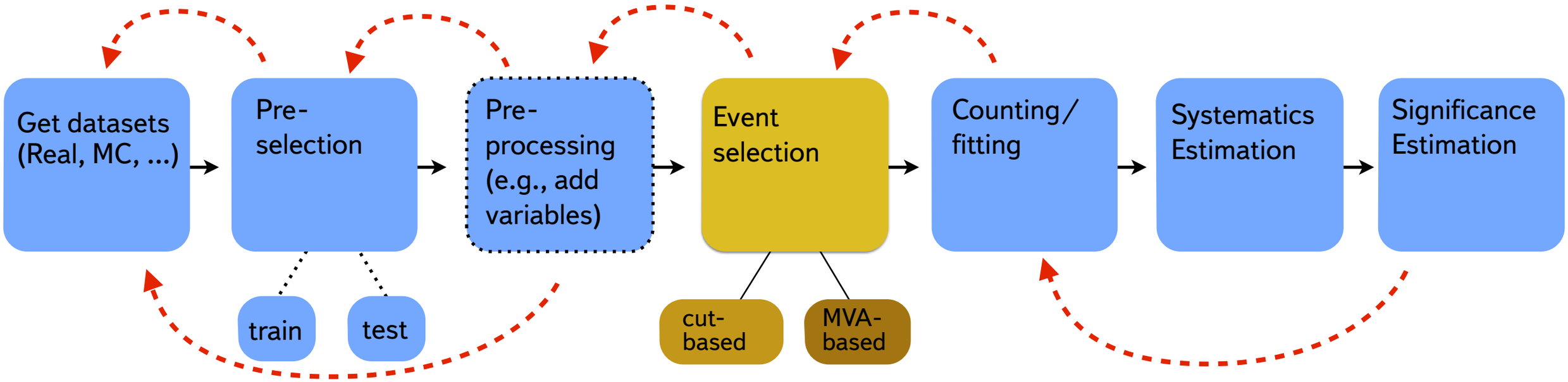
Web-based service for training MatrixNet prediction models.

Quest for sensitivity



Analysis complexity

Case: $\tau \to 3\mu$ (LHCb)



Repeat count: 10^2 10^2 10^3 10^2 10^2 10^2

Trained models: ~1500

Requires dedicated framework!

What is Event Filter going to be?

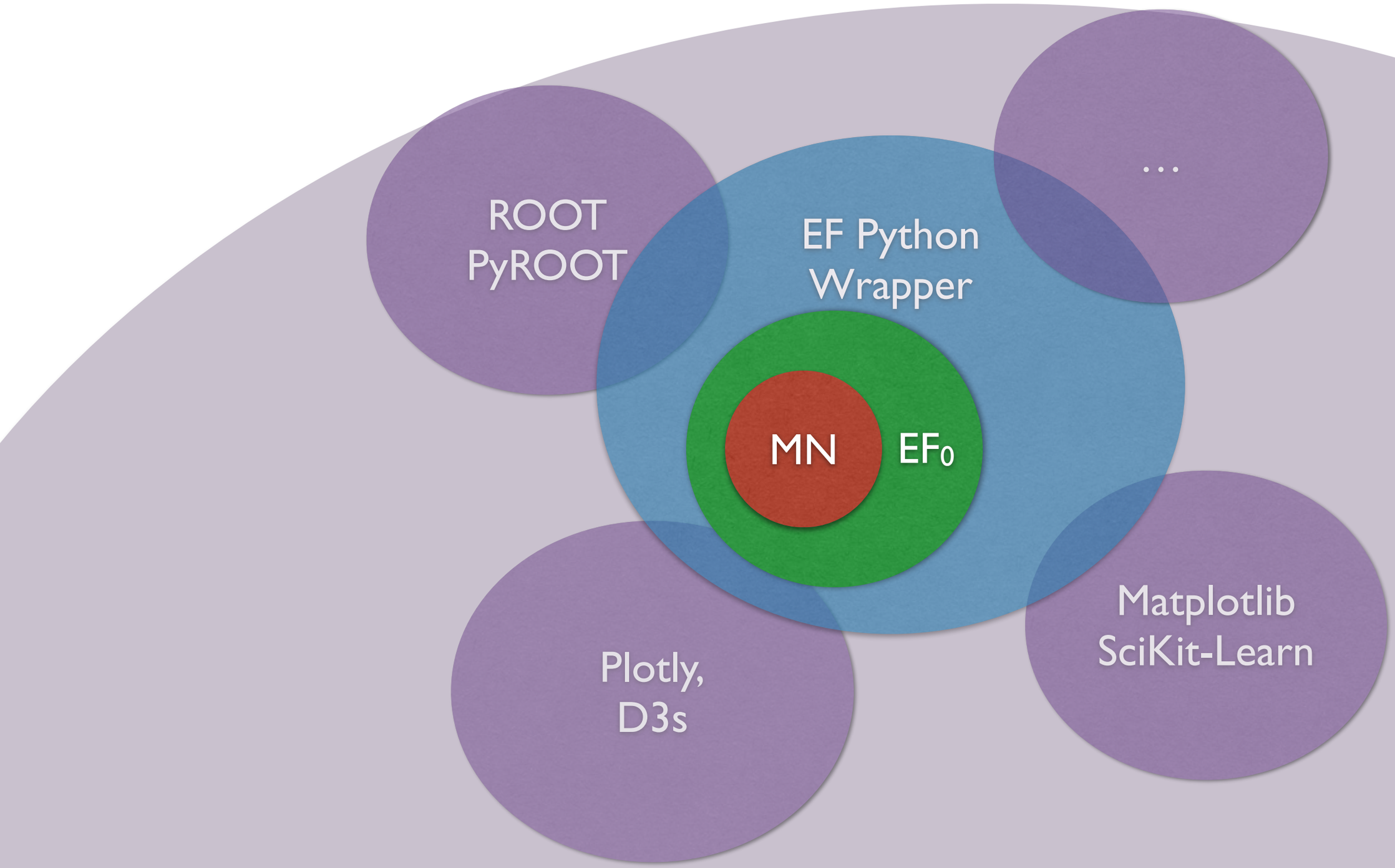
software infrastructure to support a collaborative ecosystem for computational science. It is a solution for team of researchers that allows

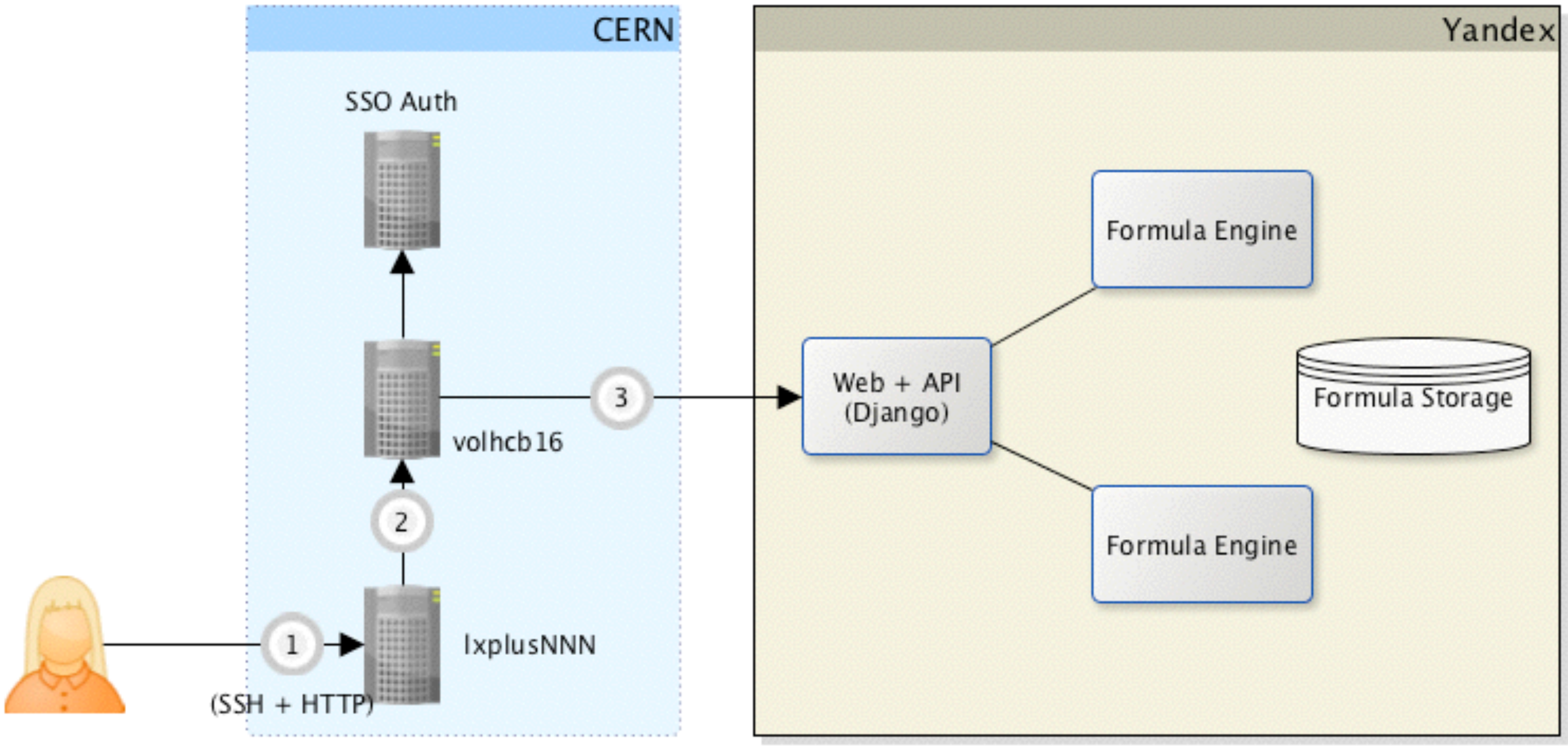
- running computational experiments on big shared datasets,
- obtaining reproducible and repeatable results,
- comparing measurable result consistently.

Event Filter features / requirements

1. research automation, i.e. defining modules that can be reused later on,
2. consistent automatic cross-check,
3. online visually enhanced shared interactive environment,
4. reproducibility of results,
5. support for standard modules,
6. scalability (performance increase as additional [hardware] resources are available),
7. [flat learning curve]

Landscape for Event Filter (EF)





EventFilter deploy landscape

«Howtos» (http://bit.ly/1ktUS4e)

- Introduction into IPython
- Event Filter (MatrixNet)
- TMVA
- Model Comparison
- Interactive Charts
- SciKit-Learn classifiers
- uBoost
- other ...

Next steps

- Testing running under CERNVM
- Clustering IPython using LSF
- Upgrade to IPython 2.0
- More analysis examples («In God we Trust, all others must bring data» (C) W.E. Deming)
- Provenance tracking (environment tracking, git)
- Running analysis jobs using modern distributed architectures (e.g. Hadoop, Impala, Drill)

Conclusion

- Event Filter development is based on real analysis needs
- Prototype for a broader analysis ecosystem. Inspired by real industry case
- Open-source, supposed to be fun
- Welcome to join! (cases? wrappers?)

anaderi@yandex-team.ru