

Yandex

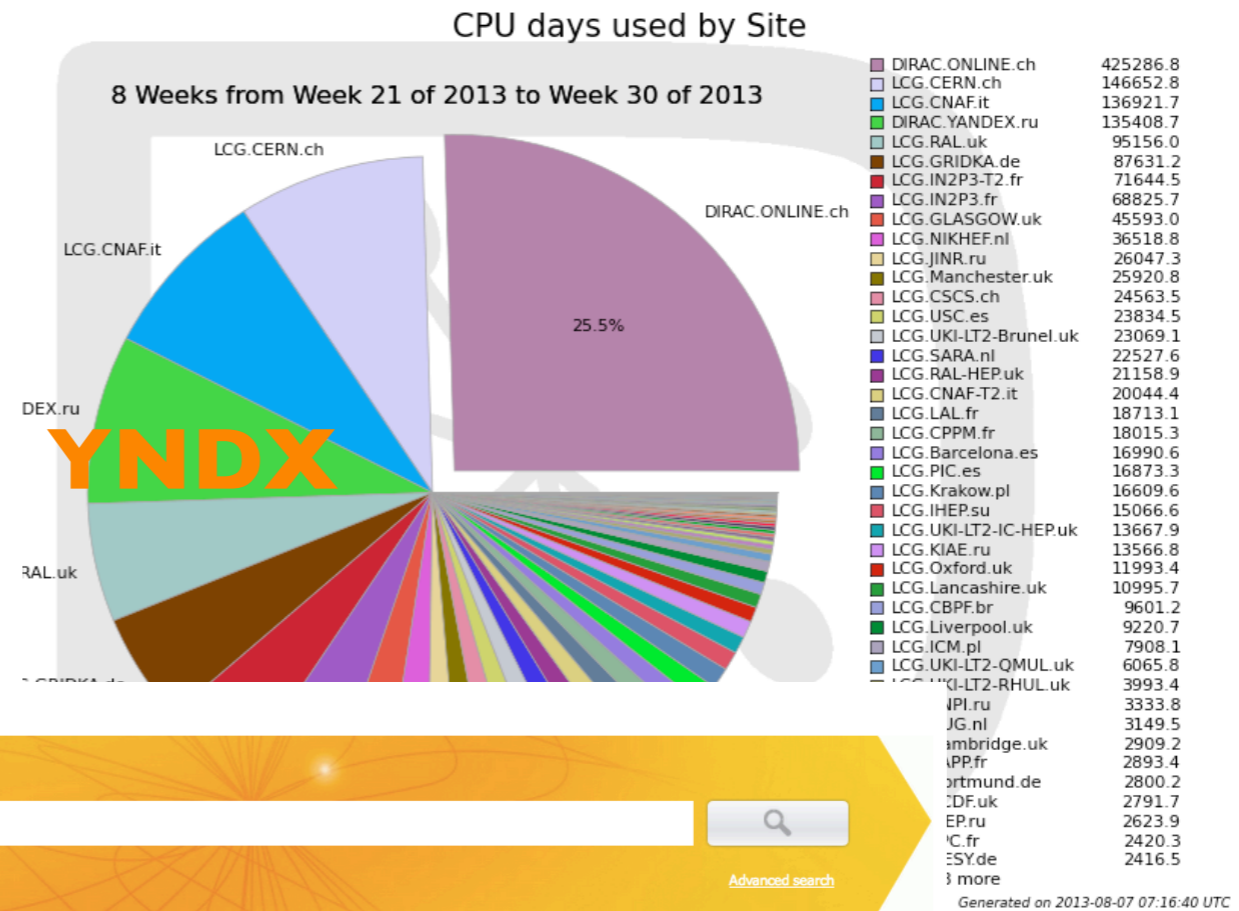


CERN openlab Projects Review

Andrey Ustyuzhanin
Head of CERN-Yandex joint projects

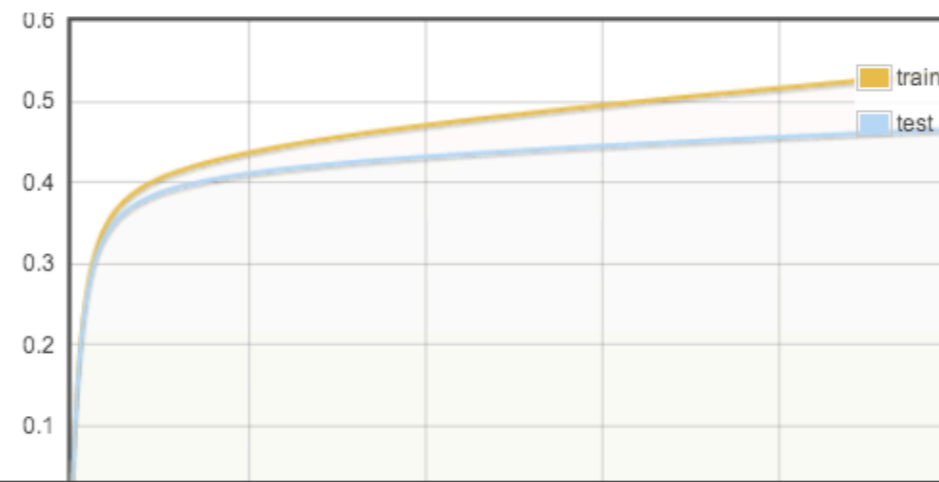
Yandex Services for CERN users

- LHCb Grid Tier 2 Cluster
- Event Index
- Event Filter



Formula «mc_Bu2X-B2psiX_2a»

Learning Factors Export



search is supported by
Yandex

Event Index



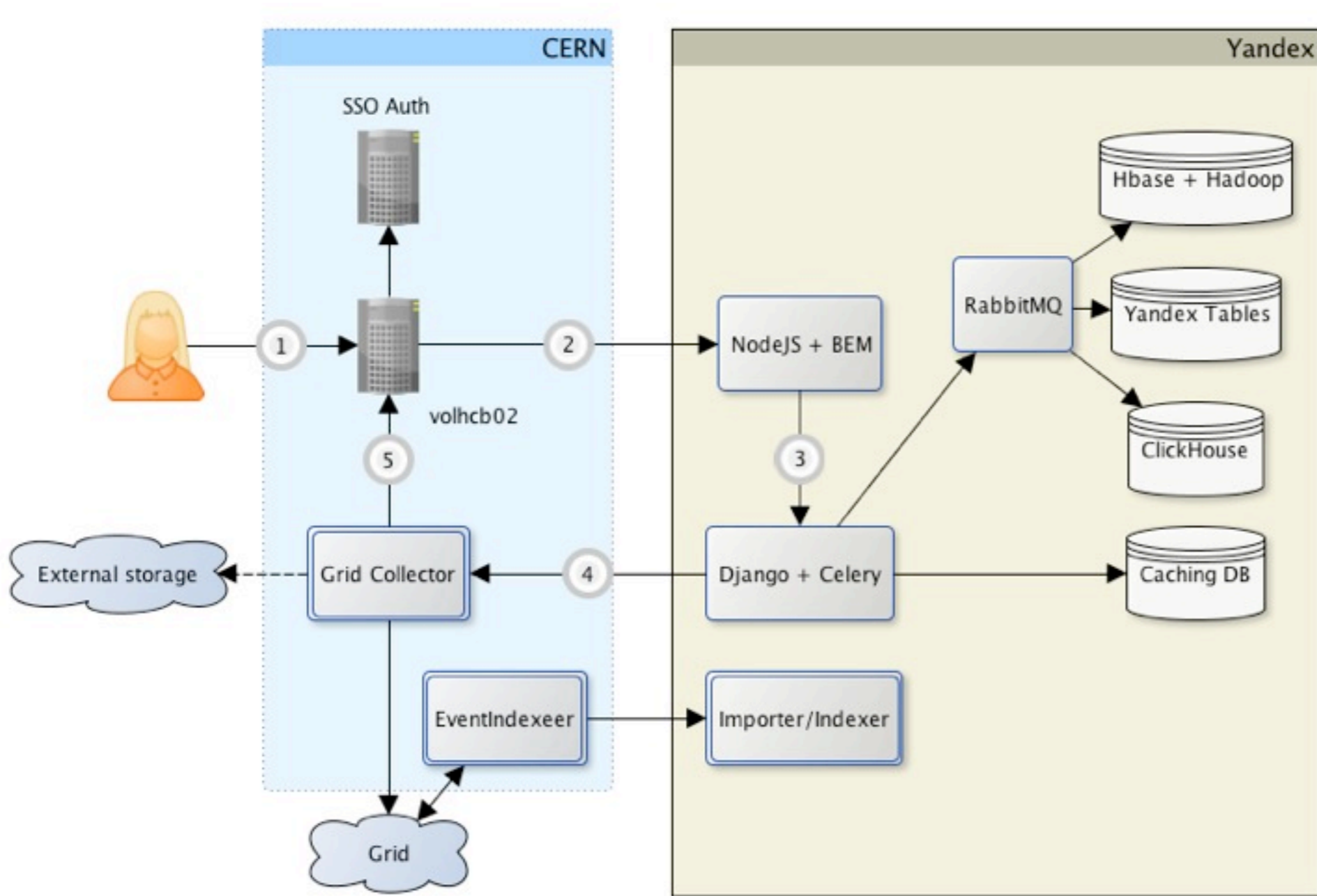
<https://eventindex.cern.ch/>
(requires LHCb membership)



Search is supported by

Yandex

Event Index Subsystems



Cluster:

8 machines,
128 CPU cores
35 TB HDD

Data per year:

~5 TB,
~ 10^{10} events,
Full scan: 20mins
index-based:
~2 seconds

Event Index

strippings_stripping_lines_dict.StrippingA1MuMuA1MuMuLineDecision=

Find

austyuzh
Last search result

[Wizard](#) [Dashboard](#)

Found 5201059 events

1 2 3 4 5 →|

[View histogram](#)

[Save search](#)

[Download](#)

Event 126577:2235701824

[details](#)

[Save to favorites](#)

Event time: 2012-08-26 12:08:58

Stripping version: 20

Stream: dimuon

Application: Brunel v43r2p2

Global activity counters

nUTClusters	-999
nVeloTracks	84
nRich2Hits	3497
nMuonTracks	6
nOTClusters	6596
nPVs	1
nLongTracks	28
nUpstreamTracks	3
nTracks	145
nTTracks	43

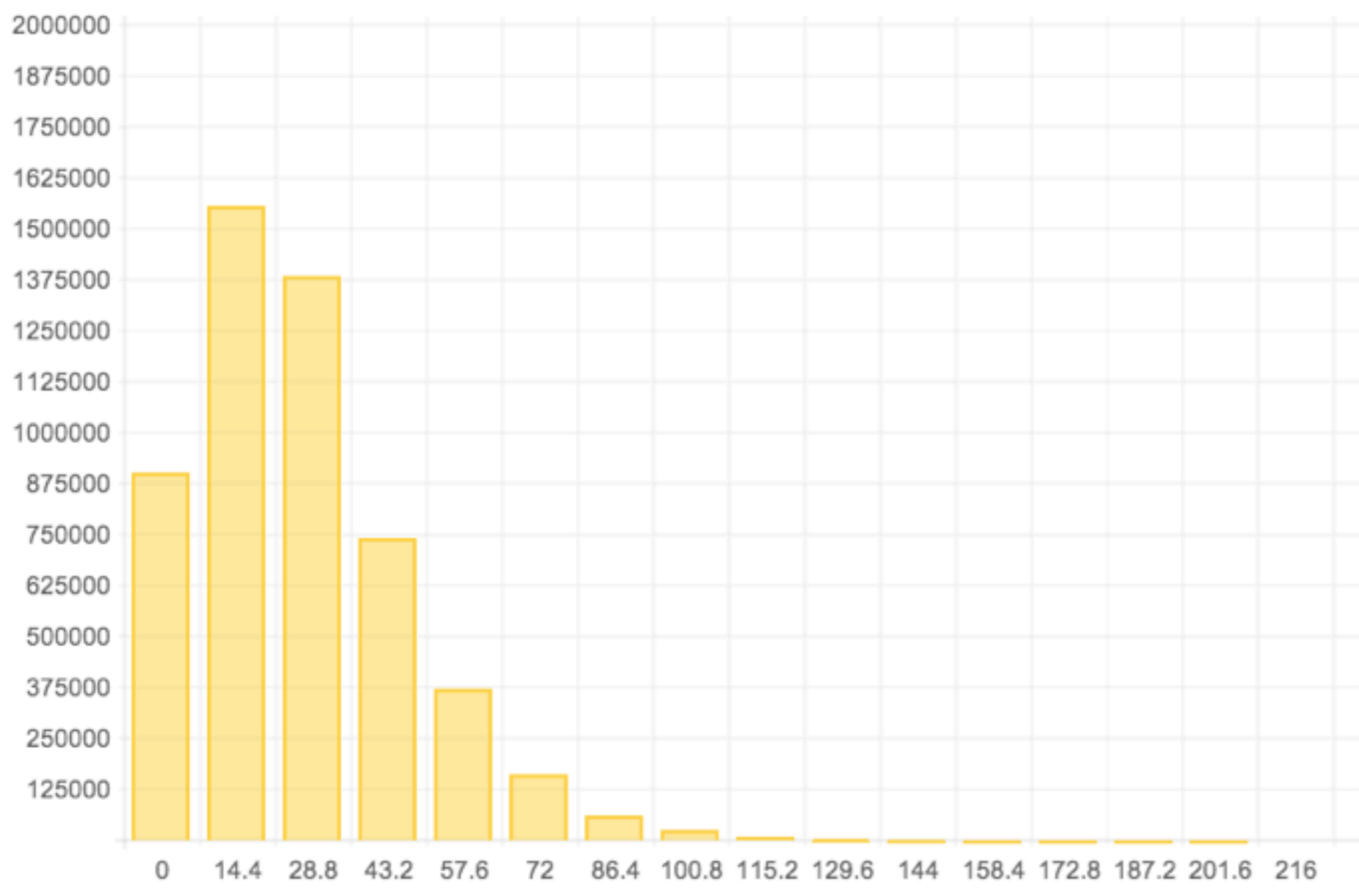
Stripping lines

StrippingStreamLeptonicDecision	0
StrippingDY2MuMuLine2Decision	1
StrippingMicroDSTDiMuonDiMuonIncLineDecision	1
StrippingA1MuMuA1MuMuLineDecision	1
StrippingStreamDimuonDecision	0
StrippingStreamEWDecision	0

Condition DB tags

DQFLAGS	dq-20120727
DDDD	1111-00100001

Wizard Dashboard



5217277 events selected

Axis X

- nBackTracks
- nDownstreamTracks
- nFTClusters
- nITClusters
- nLongTracks
- nMuonCoordsS0
- nMuonCoordsS1
- nMuonCoordsS2
- nMuonCoordsS3

Intervals count



Event Index

Find

austyuzh
Last search result

[Wizard](#) [Dashboard](#)

System status

Frontend version:
0.9.3

Backend status:
OK

Backend version:
0.008

Database version:
n/a

Searches this week / this month / overall:
1 / 31 / 31

Download requests this week / this month / overall:
0 / 2 / 2

Events count:
n/a

Users count:
4



[About](#) [Dashboard](#) [Status](#)

Search is supported by

Yandex

Search Performance Evaluation

Query	Events	Time
StrippingDstarUPB02DstDstBeauty2CharmLineDecision	7	1s
StrippingDstarUPB2DstDstKBeauty2CharmLineDecision	70	1s
Strippingb2DstarMuXKsKs_DDDDCharmFromBSemiLineDecision	1k	1s
StrippingB02DstDstKSWSLLBeauty2CharmLineDecision	11k	1s
StrippingDstarUPB2DstD0Beauty2CharmLineDecision AND stripping=20r1	38k	15s
StrippingDstarUPB2DstD0Beauty2CharmLineDecision	112k	18s
StrippingLowMultCEP_D2K3PiWS_lineDecision	214k	24s
StrippingB02DstarKDst2D0Pi_D2KSHLLBeauty2CharmLineDecision	1m	25m
StrippingXiccControlDpDecision	10m	23m
stream=ew	625m	26m 47s
stream=ew AND stripping=20 AND StrippingX2D0D0LineDecision	3m	24m 11s

Event Filter (<https://volhcb16.cern.ch/>)

Home Datasets Pools **Formulas** Predictions Tasks About

Formula	Dataset	Variables	Count	Order	Created	Score	Actions
BDT9_Flavio1_i10000_w001_x64	Flavio1	B_s0_TAU_ps, mu_MINIPS, ...	336903	1	2013-04-19 10:12:49	-0.0543496	Export, Remove
D0mumu-test							
default							
form_1							
marco_new_test							
mc_Bu2X-B2psiX_2							
mc_Bu2X-B2psiX_2a							
merger_3							
MN_13_Flavio1_i10000_w001_							
MN_13_Flavio1_i10000_w001_							

Formula «mc_Bu2X-B2psiX_2a»

Learning Factors Export

Time	Train Score	Test Score
0	0.0	0.0
1000	0.45	0.40
2000	0.48	0.43
3000	0.50	0.44
4000	0.52	0.45

Event Filter HTTP API

/api/upload_dataset

/api/train_formula

/ajax/tasks_status

/api/get_trained_formula_binary

/api/remove_dataset

...

more info at: <https://volhcb16.cern.ch/static/API.pdf>

Example:

```
anaderi@lxplus:~$ curl --digest http://volhcb16.cern.ch/api/find_pool?id=22
{
  "data": {
    "fields":
"B_DIRA_OWNPV,B_ENDVERTEX_CHI2,B_FDCHI2_OWNPV,B_IPCHI2_OWNPV,B_PT,D_ENDVERTEX_
CHI2,is_signal,Jpsi_ENDVERTEX_CHI2,Jpsi_PT,K_MINIPCHI2,K_PT,muminus_MINIPCHI2,
muplus_MINIPCHI2,pi_MINIPCHI2,pi_PT",
    "fields_count": "15",
    "id": "22",
    "name": "D0mumu_a",
    "rows_count": "53239",
    "size_on_disk": "10596801",
    "type": "CSV",
```

Event Filter API Python wrapper

Trivial Example of API usage

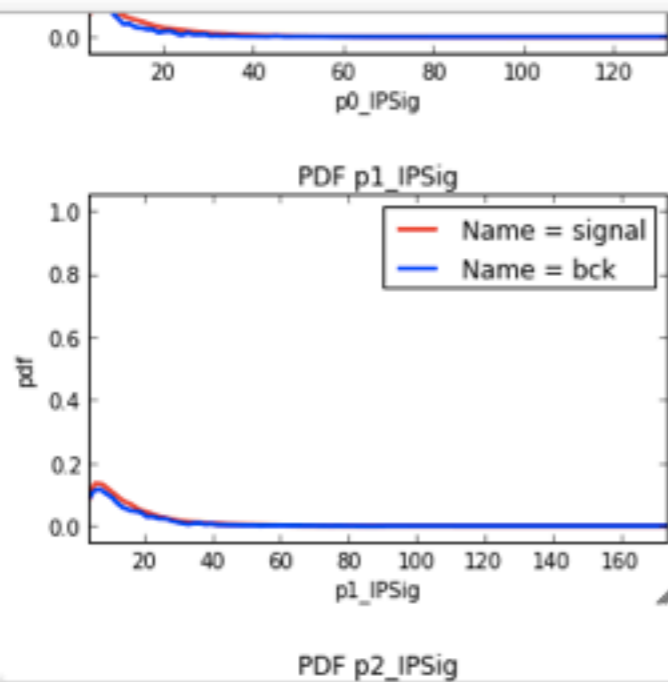
```
import Yandex_EF_API
try:
    api = Yandex_EF_API(url_base=URL, user='', password='')
    result = api.upload_dataset(train_filename, TEST_NAME)
    sleep(60) # ... or check it is ready
    result = api.train_formula(pool_id, TEST_NAME,
                              fields=[
                                  'B_DIRA_OWNPV',
                                  'B_ENDVERTEX_CHI2',
                                  'B_FDCHI2_OWNPV' ],
                              train_rows=1000)
    result = api.download_formula(
        formula_file_csv,
        formula_name=TEST_NAME,
        file_type='csv')
except Exception, e:
    print "Oops!", e
```

IPython wrapper demo

IP[y]: Notebook tau3mu_blending_all_proto_cern Last Checkpoint: Feb 12 08:43 (unsaved changes)

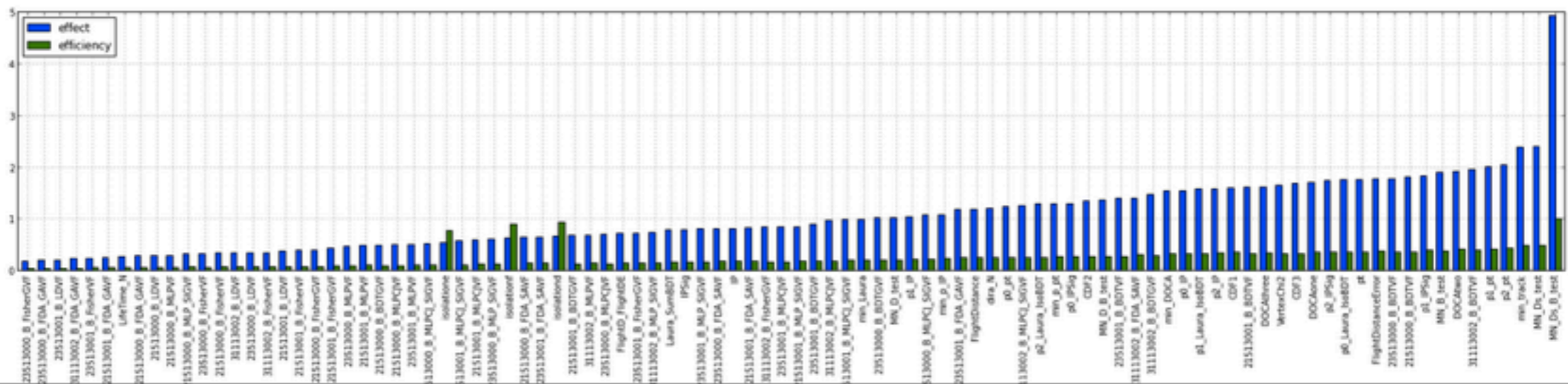
File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None



```
In [549]: importance=report['feature_importances']
importance.sort(['effect', 'efficiency'])[['effect', 'efficiency']].plot(figsize=(30,5), kind='bar')
```

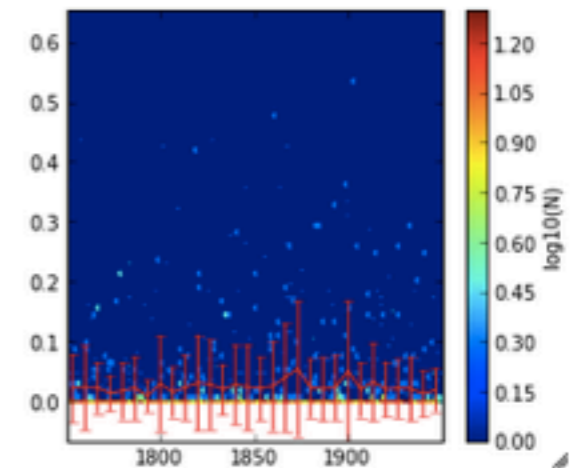
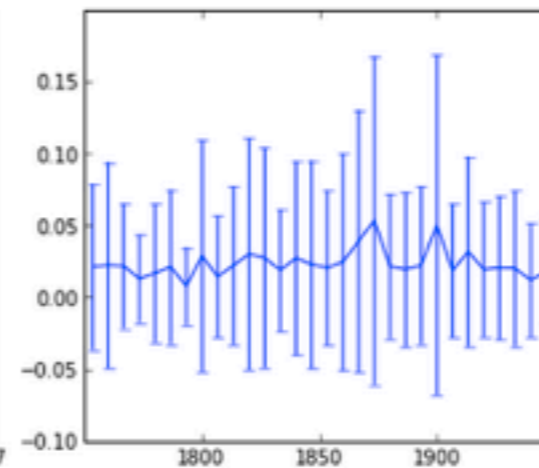
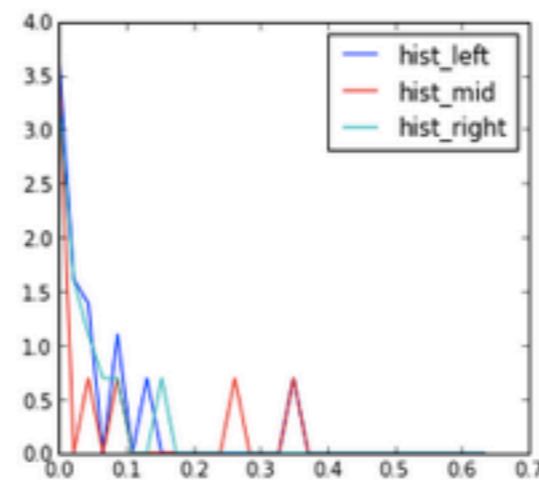
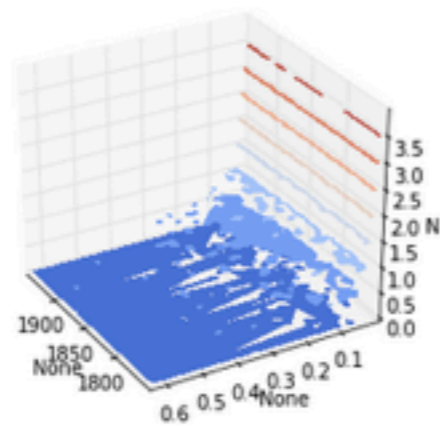
Out[549]: <matplotlib.axes.AxesSubplot at 0x2f9cddb0>



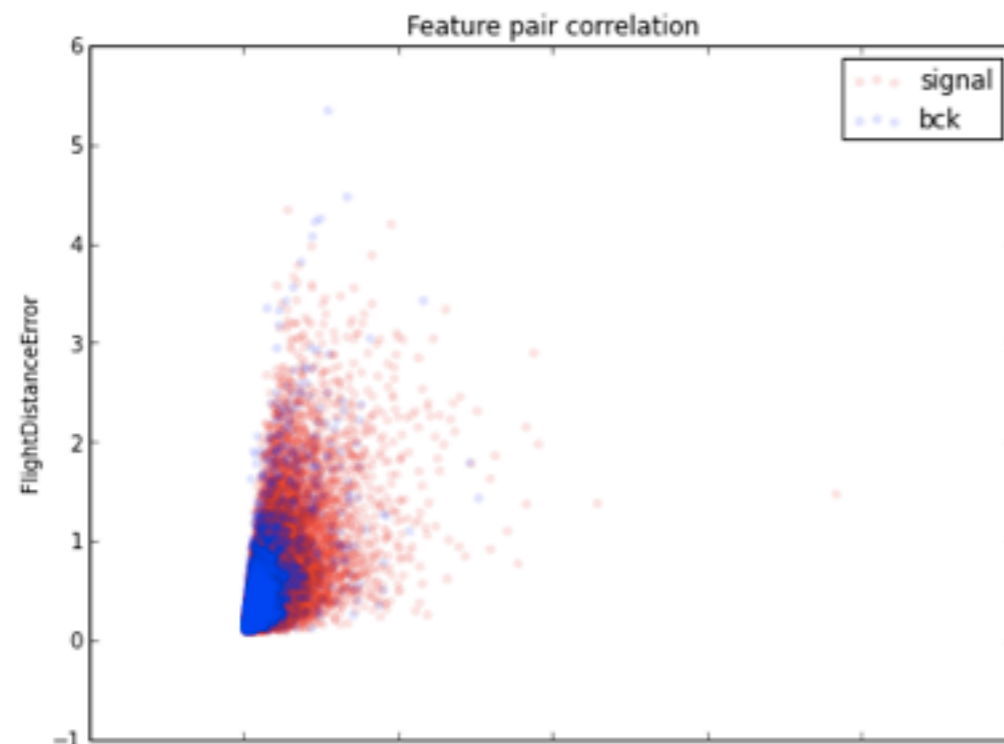
IP[y]: Notebook tau3mu_blending_all_proto_cern Last Checkpoint: Feb 12 08:43 (unsaved changes)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None



```
In [557]: for key in report['correlations'].keys():  
          Plot_Scatters(report['correlations'][key], xlabel=key[0], ylabel=key[1])
```

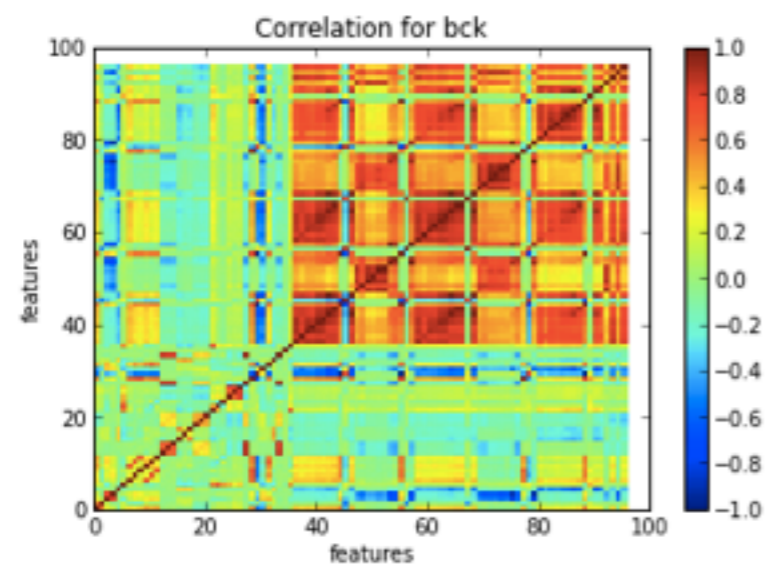
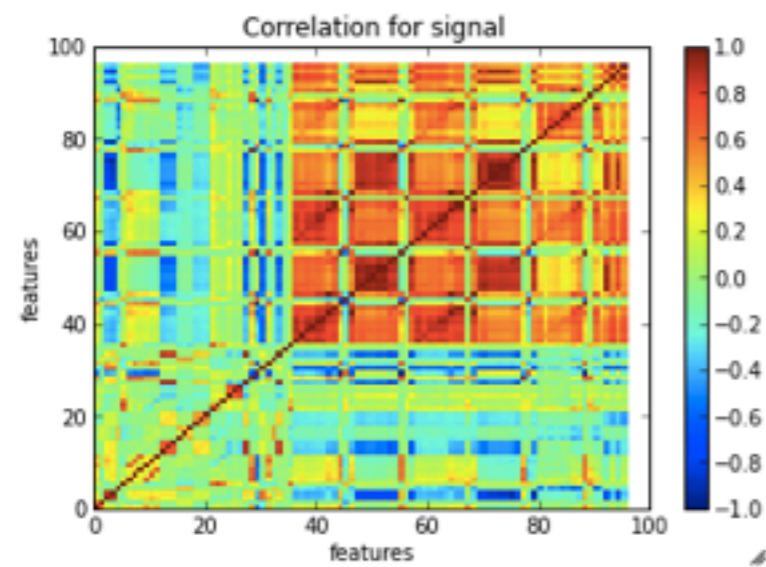


IP[y]: Notebook tau3mu_blending_all_proto_cern Last Checkpoint: Feb 12 08:43 (unsaved changes)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

```
In [547]: for key in report['feature_correlation'].keys():  
          Plot_ColorMap(report['feature_correlation'][key], figsize=(6, 4), title='Correlation for ' + key)
```



EventFilter Evaluation

$$\tau^- \rightarrow \mu^+ \mu^- \mu^-$$

J. Albrecht¹, M. Calvi^{2,3}, M. Chrzaszcz^{4,5}, L. Gavardi¹, J. Harrison⁶, B. Khanji⁴,
G. Lafferty⁶, E. Rodrigues⁶, N. Serra⁵, P. Seyfert⁷

¹ *Fakultät Physik, Technische Universität Dortmund, Germany*

² *The University of Milano Bicocca, Milano, Italy*

³ *INFN Milano Bicocca, Milano, Italy*

⁴ *Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Cracow, Poland*

⁵ *Physik-Institut der Universität Zürich, Switzerland*

⁶ *The University of Manchester, UK*

⁷ *Physikalisches Institut Heidelberg, Germany*

$$B_s \rightarrow \overline{K}^* \mu^+ \mu^-$$

M.Chrzaszcz, C.Parkinson, N.Serra,

T.Blake, C.Langenbruch

Services Roadmap

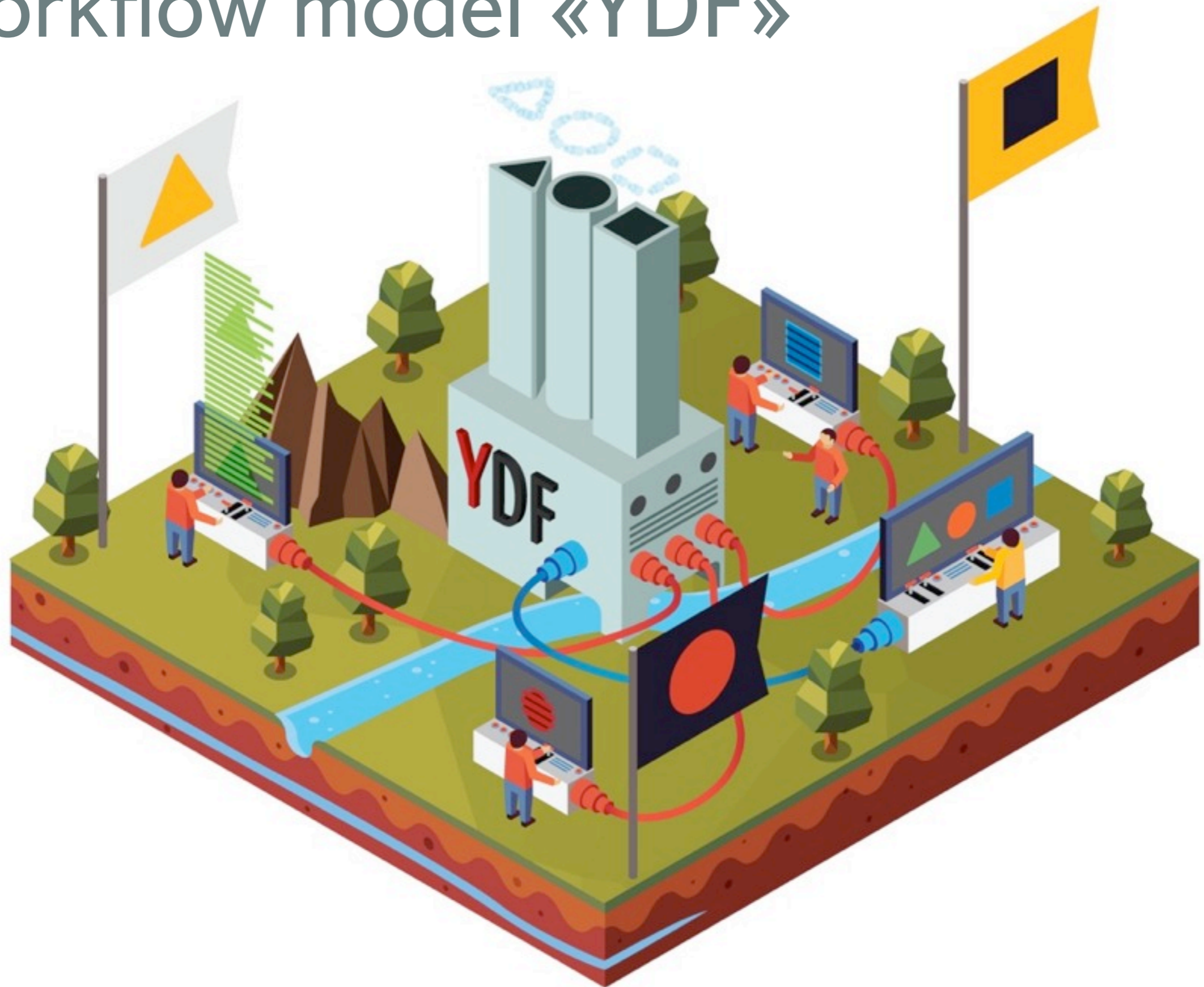
Version	Date	Event Index Features
2.0	May 2014	- Retrieval job code generation, - Summary-hash based queries
2.1	November 2014	- Histograms comparison
2.2	December 2014	Index data 2010-2011

Version	Date	Event Filter Features
2.0	May 2014	TMVA support
2.1	November 2014	Baseline training & Cross validation
2.2	December 2014	Migration to YDF

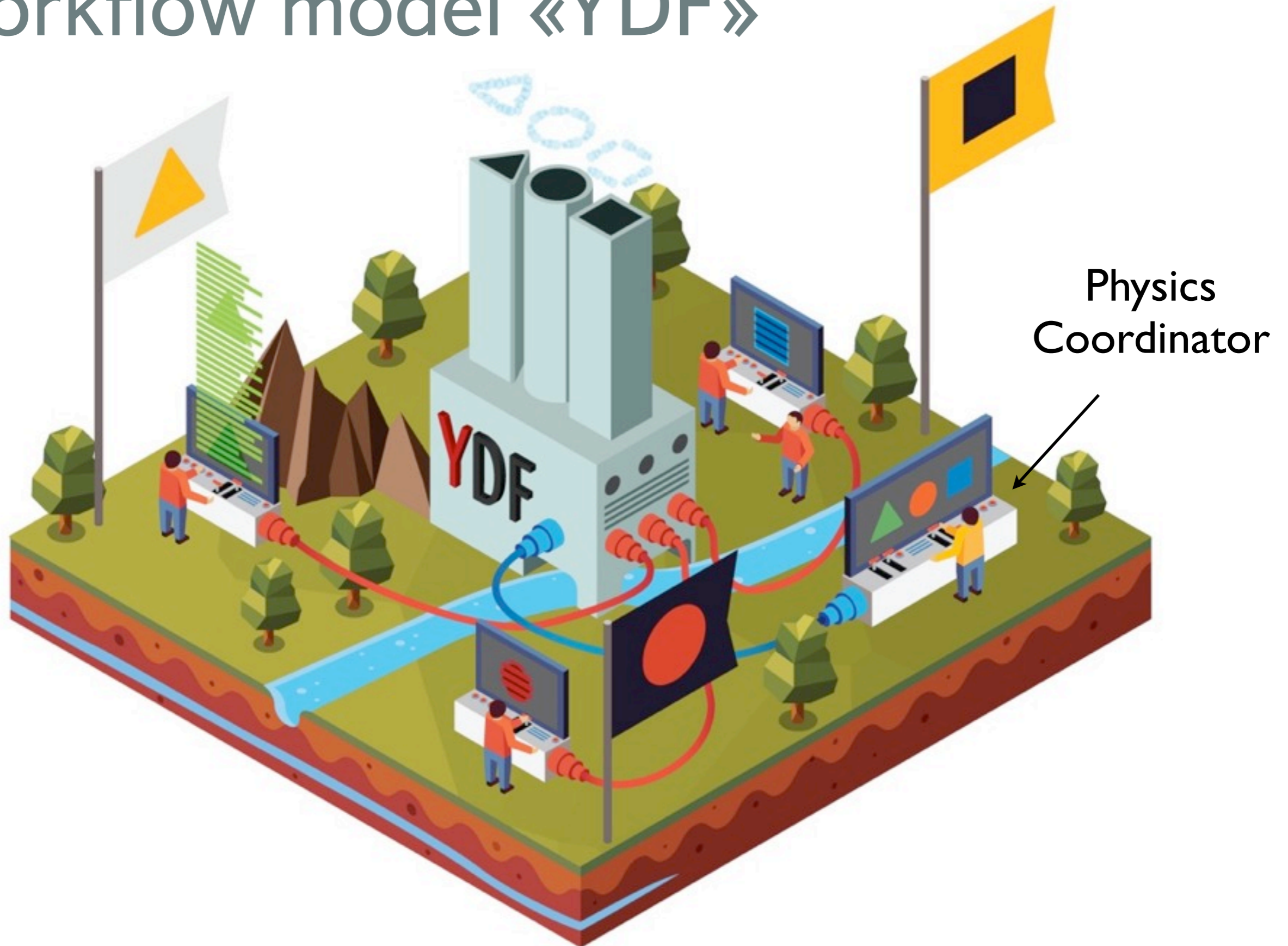
Workflow model «A»



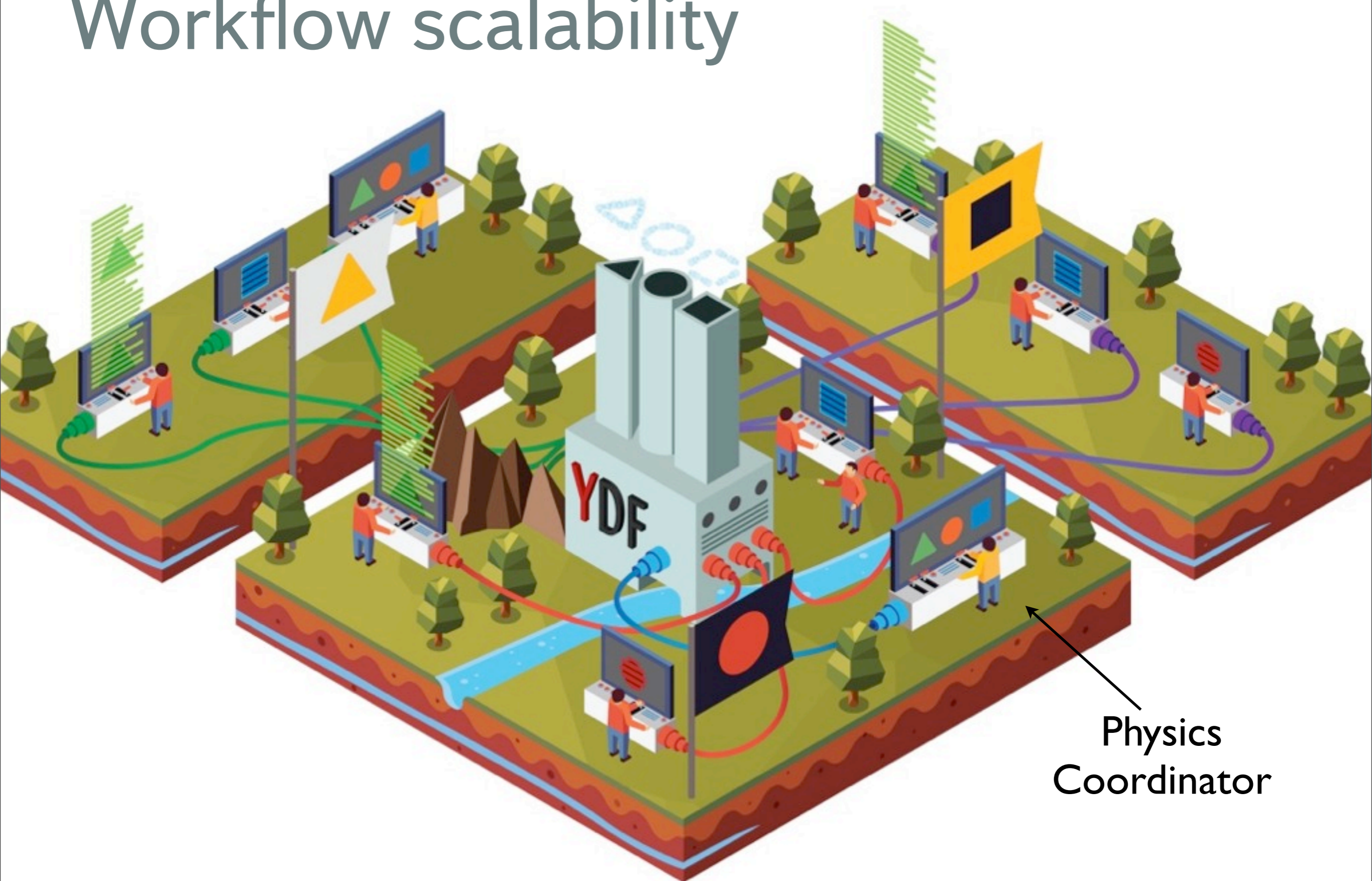
Workflow model «YDF»



Workflow model «YDF»

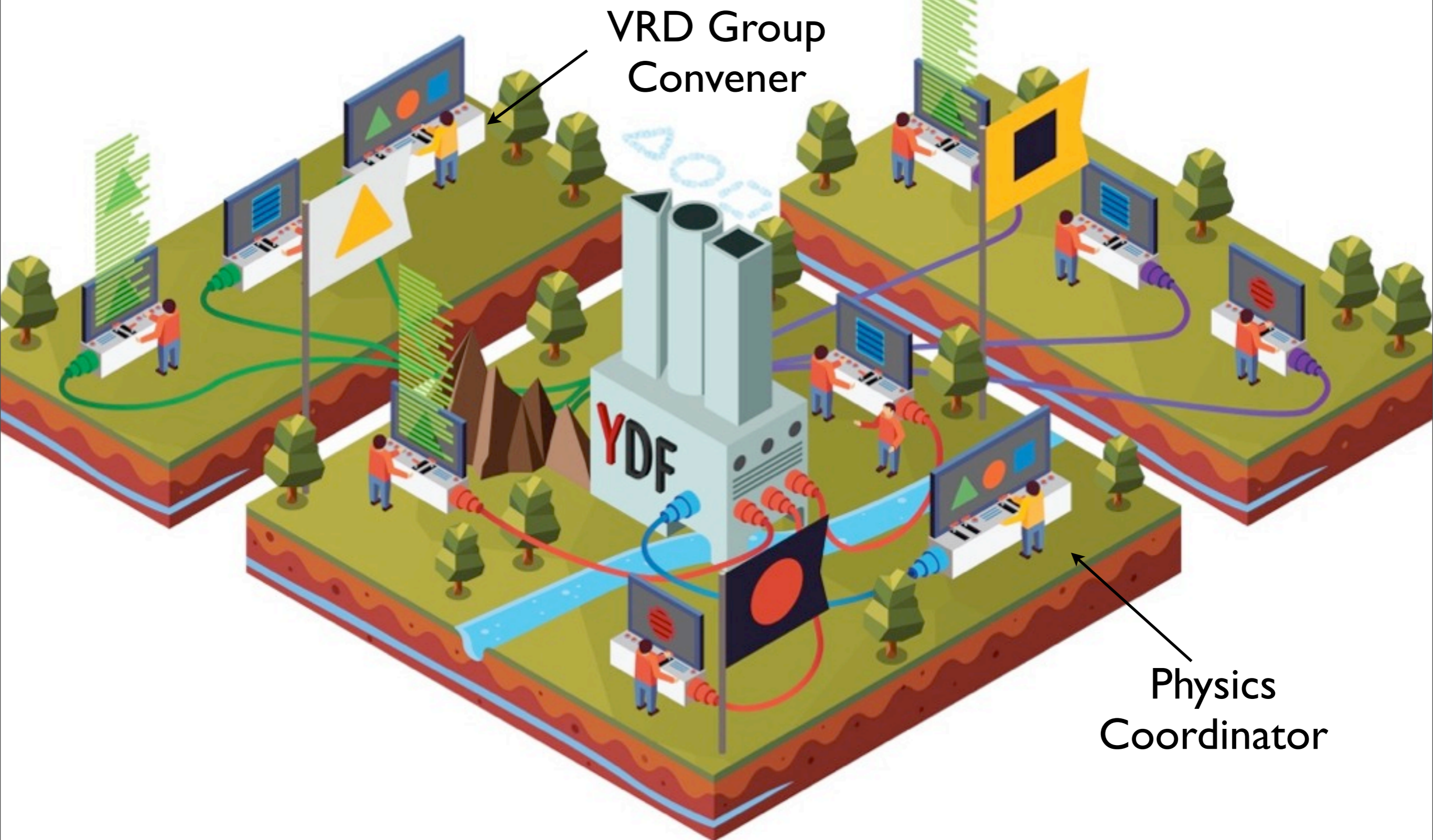


Workflow scalability



Physics
Coordinator

Workflow scalability



Conclusion

› EventIndex update:

- indexed 2012 data (5TB, 10^{10} events)
- added support for histograms

› EventFilter update:

- used in $\tau \rightarrow 3\mu$ and $B_s \rightarrow \overline{K}^* \mu^+ \mu^-$ decay analysis
- added IPython wrappers to evaluate trained model quality

› Roadmap for 2014 presented