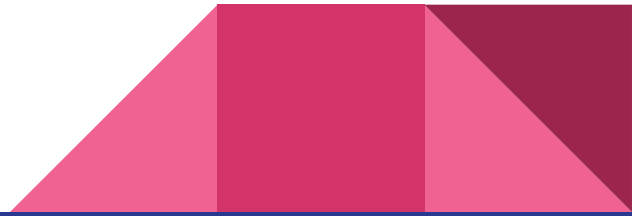
Integrating REANA into ROB

Ajay R. Rawat
IRIS-HEP fellow
University of Washington, Seattle

Reproducible Open Benchmarks for Data Analysis Platform (ROB)

- Controlled competition-style environment platform
- Evaluate data analysis workflows
- Benchmark different workflows and rank them
- Reproduce the results of a workflow



Inspiration for ROB

- Top Tagger comparison (https://ar

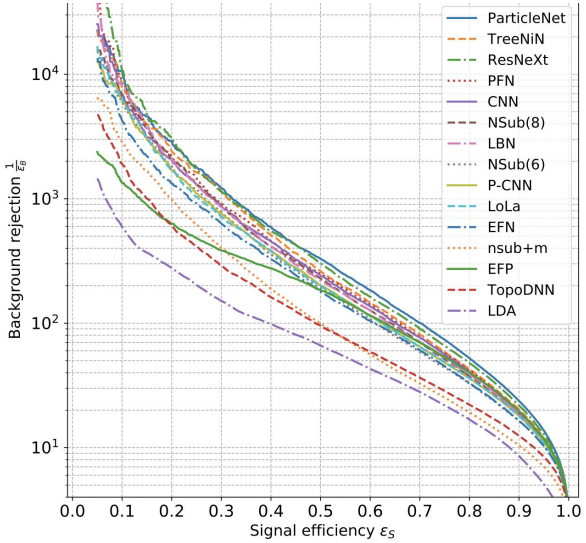
Task: Classify the Top Quark jets

From the background QCD jets.

Benchmarks:

- AUC of ROC curve
- Accuracy
- (and Rejection power [not shown here])

	AUC	Acc
CNN [16]	0.981	0.930
ResNeXt [31]	0.984	0.936
TopoDNN [18]	0.972	0.916
Multi-body N-subjettiness 6 [24]	0.979	0.922
Multi-body N-subjettiness 8 [24]	0.981	0.929
TreeNiN [43]	0.982	0.933
P-CNN	0.980	0.930
ParticleNet [47]	0.985	0.938
LBN [19]	0.981	0.931
LoLa [22]	0.980	0.929
LDA [54]	0.955	0.892
Energy Flow Polynomials [21]	0.980	0.932
Energy Flow Network [23]	0.979	0.927
Particle Flow Network [23]	0.982	0.932
GoaT	0.985	0.939



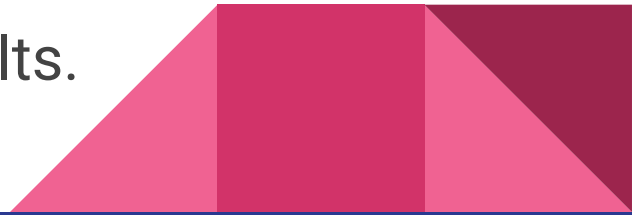
ROB Goals

ROB is designed to expand the approach used in
Top Tagger comparison and to generalise it for all data analysis workflows

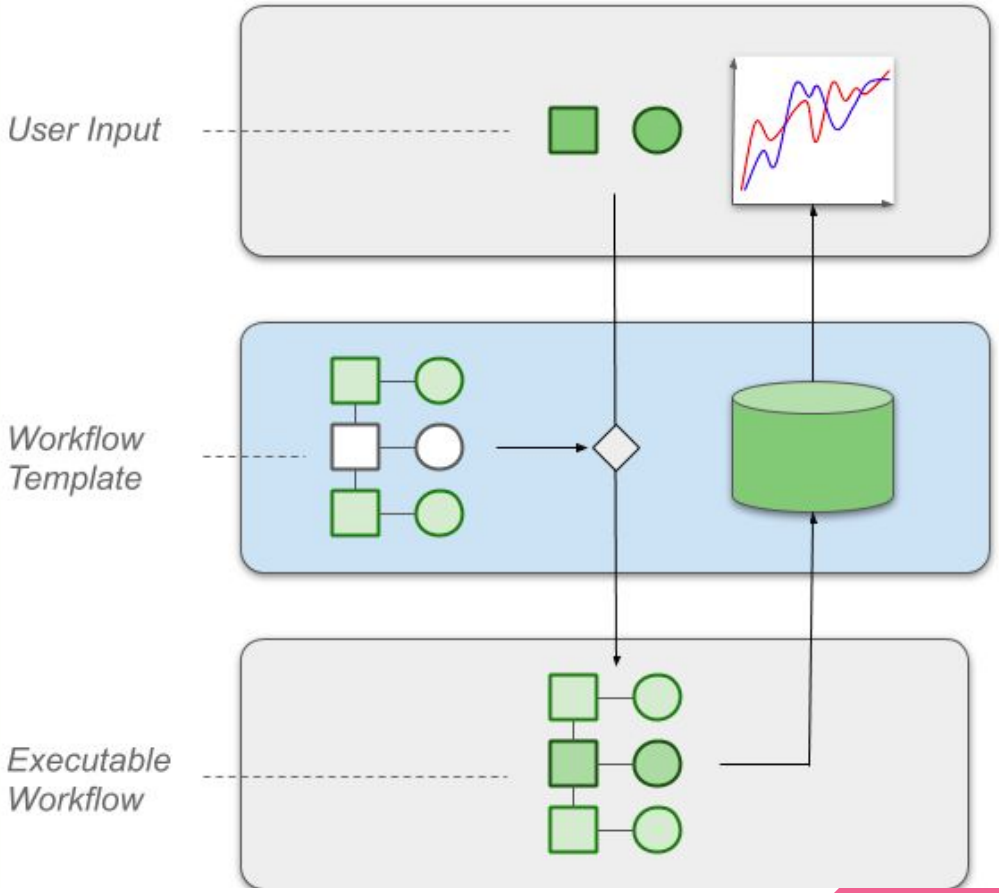
Goals:

(1) reduce the amount of time required to organize and evaluate such benchmarks

(2) ensure reproducibility of benchmark results.



ROB:



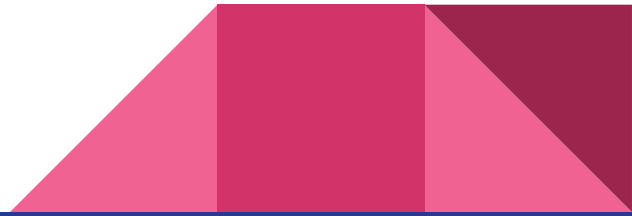
ROB
Reproducible Open Benchmarks

flowServ

Back-End

3 layers

- User Interface (Command Line Client, Web User Interface)
- Flowserv
- Workflow Engine: (local or cloud like REANA/AWS)

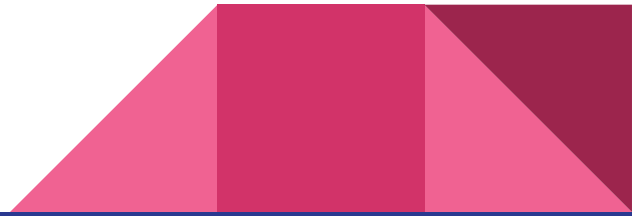


Benchmarking

- Create a task and the dataset
- Users submit their workflow and code
- ROB runs the workflow on the specified engine
- Benchmarks such as model accuracy/ AUC of ROC curve are obtained using the testing data

Benchmarking (contd.)

- Users can provide environments like Docker to run their workflow
- All of the workflows are then ranked based on these benchmarks
- They are evaluated on the same dataset



REANA

What is REANA?

REANA is a **re**producible **ana**lysis platform allowing scientists to run containerised data analysis pipelines on remote compute clouds.

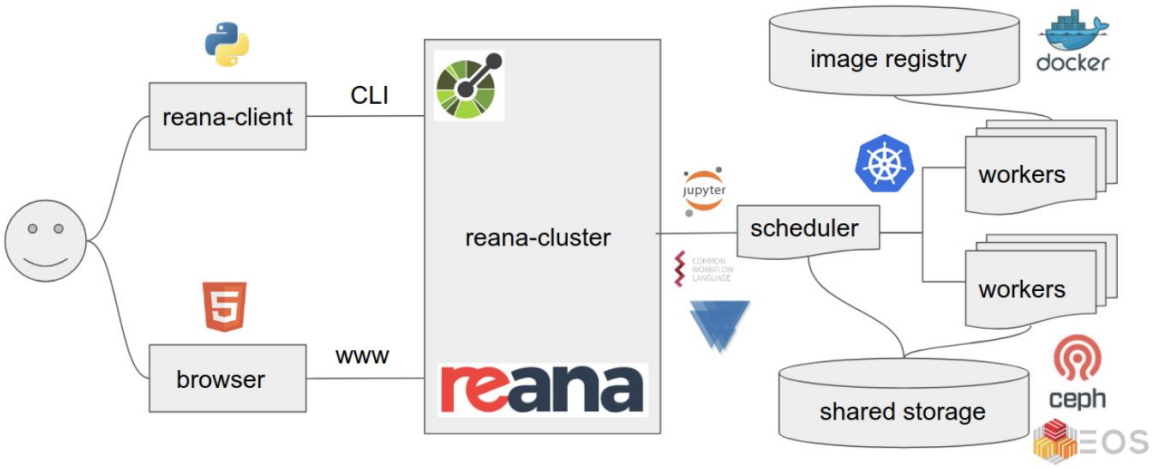


Image from:

http://docs.reana.io

REANA workflows:

REANA workflows have mainly 3 sections:

- Inputs:
  - Input files
  - parameters
- Workflow:
- Commands to run this workflow
- Outputs (files) produced by the workflow

```yaml
version: 0.3.0
inputs:
  files:
    - code/helloworld.py
    - data/names.txt
  parameters:
    helloworld: code/helloworld.py
    inputfile: data/names.txt
    outputfile: results/greetings.txt
    sleeptime: 0
workflow:
  type: serial
  specification:
    steps:
      - environment: 'python:2.7-slim'
        commands:
          - python "${helloworld}"
              --inputfile "${inputfile}"
              --outputfile "${outputfile}"
              --sleeptime ${sleeptime}
outputs:
  files:
   - results/greetings.txt
```

REANA hello world YAML file

Using a REANA cluster

- Use an existing REANA cluster eg. at CERN
  - If you have a cern account you can get an access token

- Deploy a REANA cluster on your local machine/server

Running workflows on REANA

- Command Line Interface **reana-cleint**
- Python Package

Your REANA token

In order to use your token, make sure you have reana-client installed and run:

```
$ export REANA_SERVER_URL=https://reana.cern.ch
$ export REANA_ACCESS_TOKEN=[redacted]
```

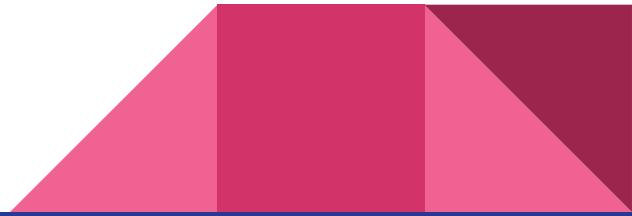
Integrating REANA in ROB

For ROB to run workflows on REANA, it uses the REANA python package to:

- Create a workflow on REANA by converting the YAML file into JSON
- Uploading all the input files to REANA server
- Starting the Workflow
- Monitoring the status of the workflow at certain time intervals
- Downloading the results once the workflow has finished execution

Benefits of integrating REANA in ROB

- All the required information to reproduce a workflow is stored on the RENA server
- REANA server also stores the logs and output files
- ROB can now run computationally heavy workflows on specific REANA clusters
- ROB can now run multiple submissions at a time on REANA and then rank benchmarks locally



Sources

- https://arxiv.org/abs/1902.09914
- https://github.com/scailfin/rob-client
- https://github.com/scailfin/rob-ui
- http://docs.reana.io