



A prominent researcher left our group, and a new postdoc has arrived, keen to advance their work. However, reproducing the previous research has proven difficult.

You plan to present these results at a plenary session in two weeks. Hopefully, the excellent results you've achieved yourself will suffice. These were obtained a few months ago, and you've encountered difficulties in reproducing them.

The problem is complicated by the fact that the machine you were using was updated, resulting in significantly slower experiment run times. Additionally, reconfiguring the proper environment has proven to be a challenge.



Denis Boyda ✓

ML for Science: Best Practices

MLOps

*I hope to propose some methods that could assist you in conducting your ML4Science research **reliably, efficiently, and consistently.***

Denis Boyda, IAIFI/CTP
February 02, 2024

Every time you run an ML experiment, you should ensure that all hyperparameters are captured and saved. The aim is to create a scenario where **it's impossible for any hyperparameter not to be saved.**

```
1 import torch
2 import logging
3 from sacred import Experiment
4
5
6 logging.basicConfig(level=logging.CRITICAL)
7 ex = Experiment('hello_config', save_git_info=False)
8
9
10 @ex.config
11 def my_config():
12     opt_type = 'Adam'
13     opt_cfg = {'lr': 0.1}
14
15
16 @ex.automain
17 def my_main(_config):
18     model = torch.nn.Linear(1, 1)
19     optClass = getattr(torch.optim, _config['opt_type'])
20     optimizer = optClass(model.parameters(), **_config['opt_cfg'])
21     print(optimizer)
22
```

Sacred

*Every experiment is sacred
Every experiment is great
If an experiment is wasted
God gets quite irate*

```
(torch-02-22)boyda@platiypus:~/codes/sacred$
> python exp6.py --force with opt_type=LFBGS \
> opt_cfg="{ 'lr':1e-5, 'max_eval': 25}"
LFBGS (
Parameter Group 0
  history_size: 100
  line_search_fn: None
  lr: 1e-05
  max_eval: 25
  max_iter: 20
  tolerance_change: 1e-09
  tolerance_grad: 1e-07
)
```

my_app.py

```
from omegaconf import DictConfig, OmegaConf
import hydra

@hydra.main(version_base=None)
def my_app(cfg: DictConfig) -> None:
    print(OmegaConf.to_yaml(cfg))

if __name__ == "__main__":
    my_app()
```

```
$ python my_app.py +db.driver=mysql +db.user=omry +db.password=secret
db:
  driver: mysql
  user: omry
  password: secret
```



Hydra

Code changes are always saved



Weights & Biases Docs

Details for: `pt-cnn-hier-v0-beta5.5` (Id: 4417) ↗

Metrics Plot

Captured Out

Experiment Details

Host Info

Run Info

Meta Info

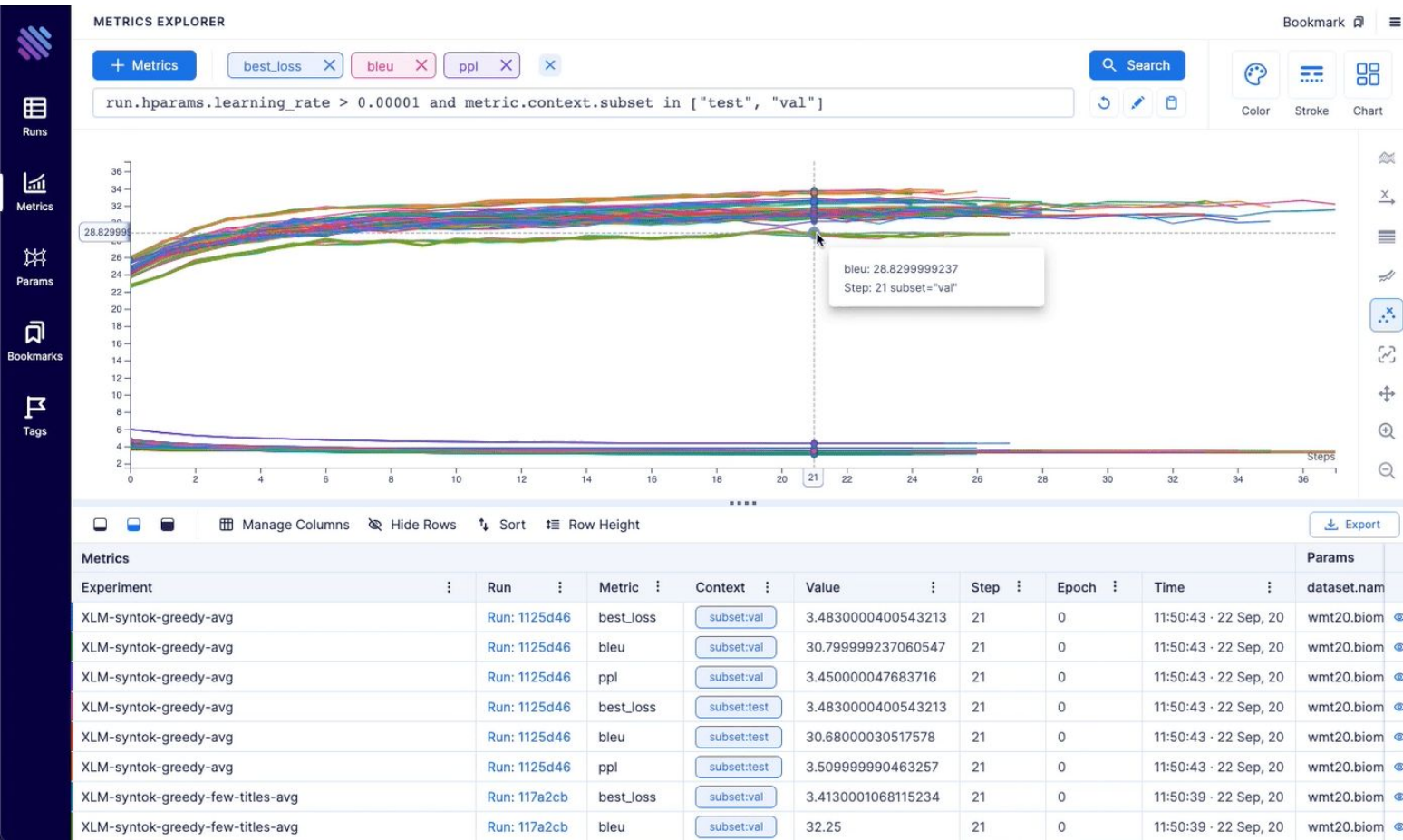
```
"root" : { 6 items
  "name" : "pt-cnn-hier-v0-beta5.5"
  "base_dir" : "/home/Lqcd/boyda/codes/lqft-flow-2"
  "sources" : [...] 8 items
  "dependencies" : [...] 3 items
  "repositories" : [ 8 items
    0 : { 3 items
      "url" : "git@github.com:"
      "commit" : "797c243a4e9574d6ba73b826cd692d4a7b0fc013"
      "dirty" : true
```

Sacred

*Every experiment is sacred
Every experiment is great
If an experiment is wasted
God gets quite irate*

copper-flower-115 ▾	vocal-cherry-116 ▾
Expand 1 lines ...	
2 <code>import time</code>	2 <code>import time</code>
3 <code>from IPython.core.display import display, HTML</code>	3 <code>from IPython.core.display import display, HTML</code>
4 <code>import matplotlib.pyplot as plt</code>	4 <code>import matplotlib.pyplot as plt</code>
5 - <code>wandb.init(project="nbviewer")</code>	5 + <code>wandb.init(project="nbviewer" , dir="/tmp")</code>
6	6
7 <code>display(HTML("COOL"))</code>	7 <code>display(HTML("COOL"))</code>
8	8
Expand 13 lines ...	

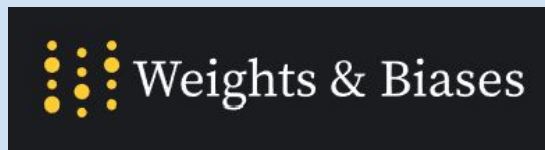
A convenient dashboard is always at your fingertips. Track and analyze your data with ease, share it with collaborators, and work together seamlessly.



MLOps in a nutshell

MLOps is a paradigm, including aspects like best practices, sets of concepts, as well as a development culture when it comes to the end-to-end conceptualization, implementation, monitoring, deployment, and scalability of machine learning products.

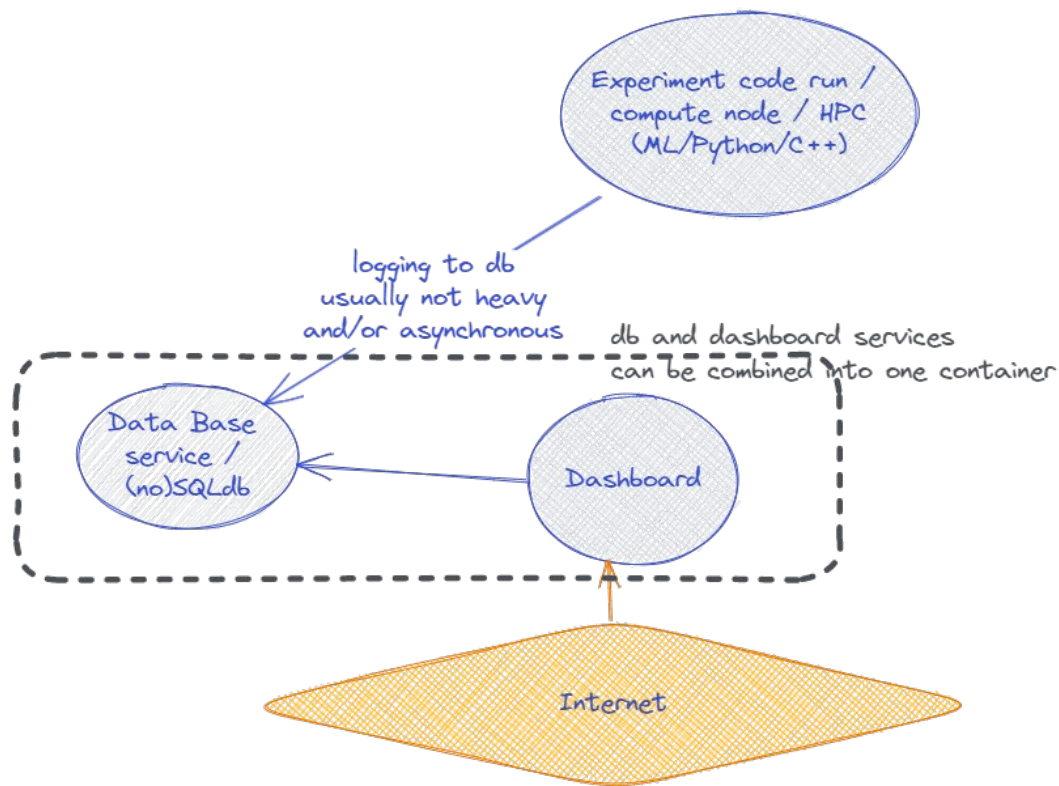
- Hyperparameters/configuration tracking
- Live information (stdout, stderr, results)
- Artifacts (models, datasets) control and versioning
- Code control and versioning
- Environment configuration
- Fail trace
- Dashboard
- Database with API



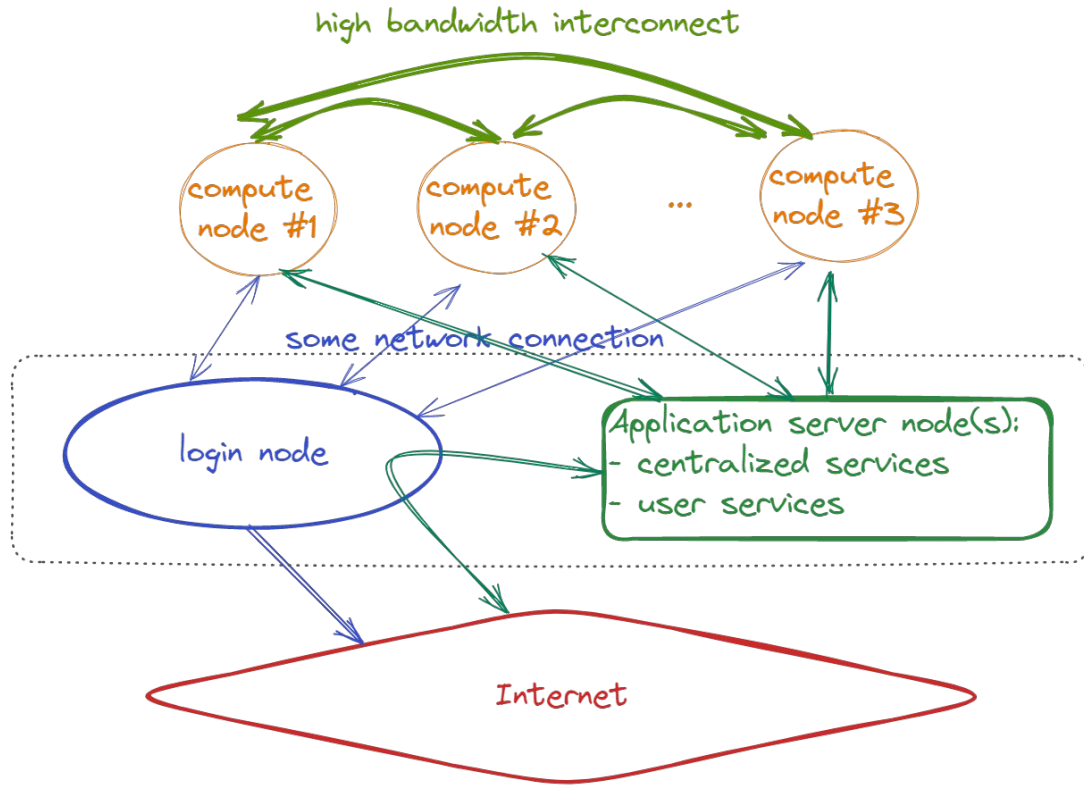
Sacred

*Every experiment is sacred
Every experiment is great
If an experiment is wasted
God gets quite irate*

MLOps Organization Overview

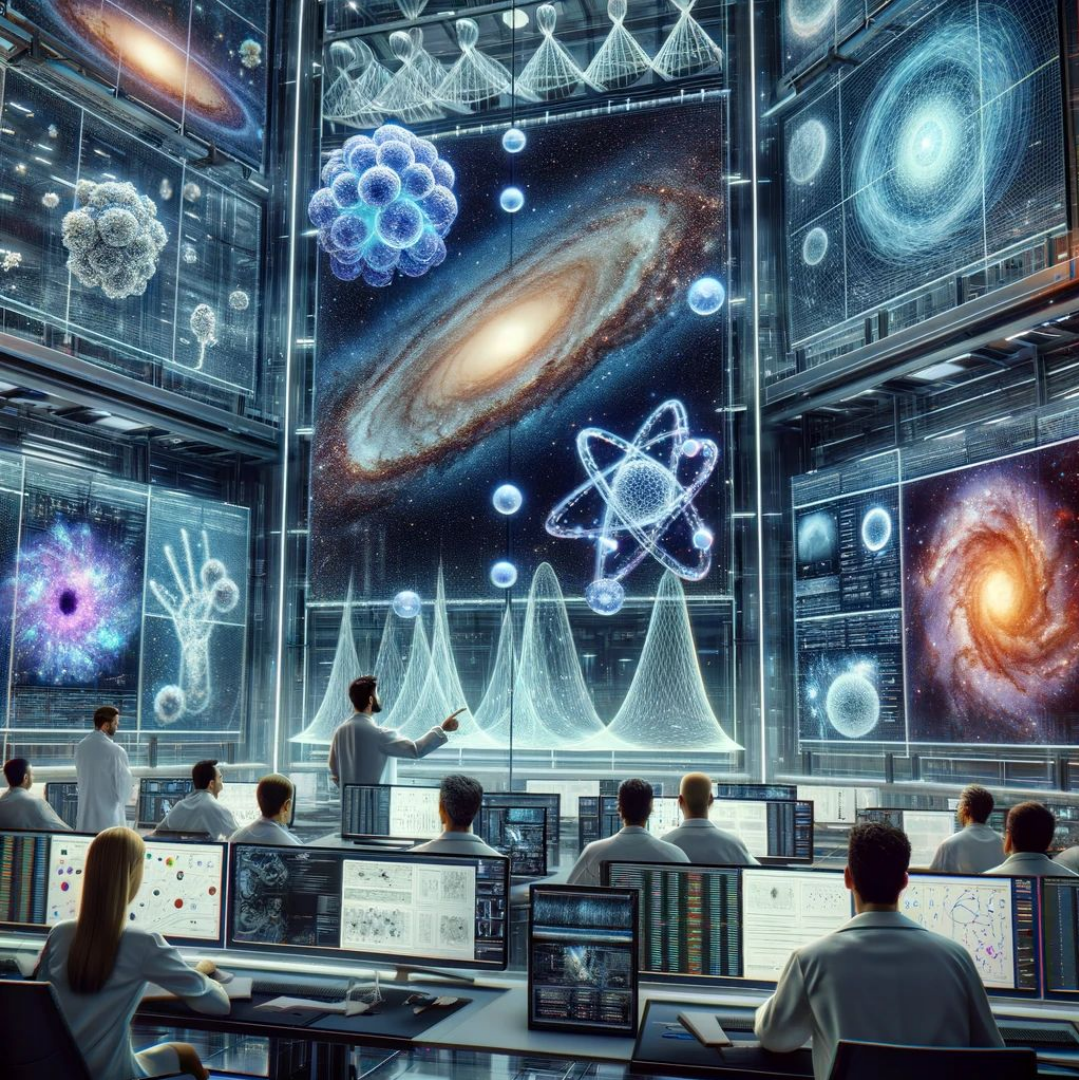


Why subMIT?



Main difference between compute nodes and application server is

- users use **compute nodes** with fixed resources for a fixed time
- at **application server** services takes small resources but all the time



Thank you for your attention.

I'm really excited to serve community and hope you picked up some useful tips from my talk.

If you're into HPC and scaling ML for science, I'd love to chat more about it. Don't hesitate to hit me up!