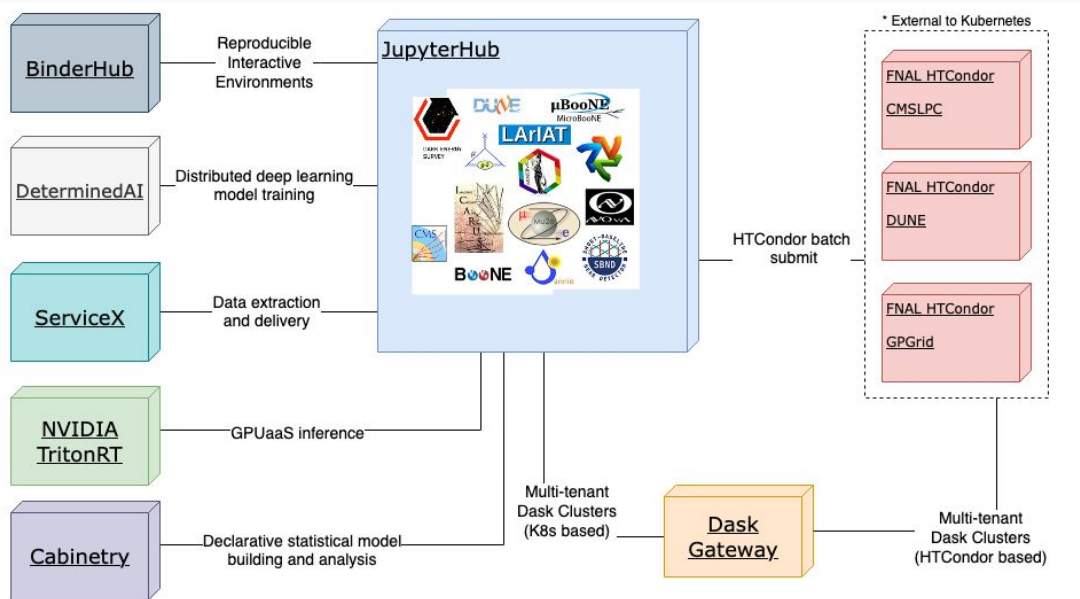


# Training NuGraph on the EAF

Maggie Voetberg  
NuGraph All Hands July 2024



# Background on the Elastic Analysis Facility



- JupyterHub based service - python native
- Request personal server attached to resources
- Connected to Ceph file system for data access

# In their own words:




The **Fermilab Elastic Analysis Facility** is a kubernetes-based platform providing services for integrated data, software and computational resources to execute one or more elements of scientific analysis workflows for Fermilab experiments and scientists. These resources are centrally managed by the Scientific Computing Division at Fermilab as part of the EAF project and generally available for all Fermilab users.



# Getting Access

- Comes free with your Fermi user account
- Requires VPN or Proxy Access
- Different servers require different permissions, everyone has access to the generic version
- If you were a part of fwk on Wilson cluster, you should be able to get into FIFE now



 Fermilab  
Generic SL7 notebooks

**CPU Interactives**

AL9

**NVIDIA® A100 GPU**

AL9 - 10GB GPU slot

**Rescue Image**

Rescue Image - for use if overquota or other startup issues

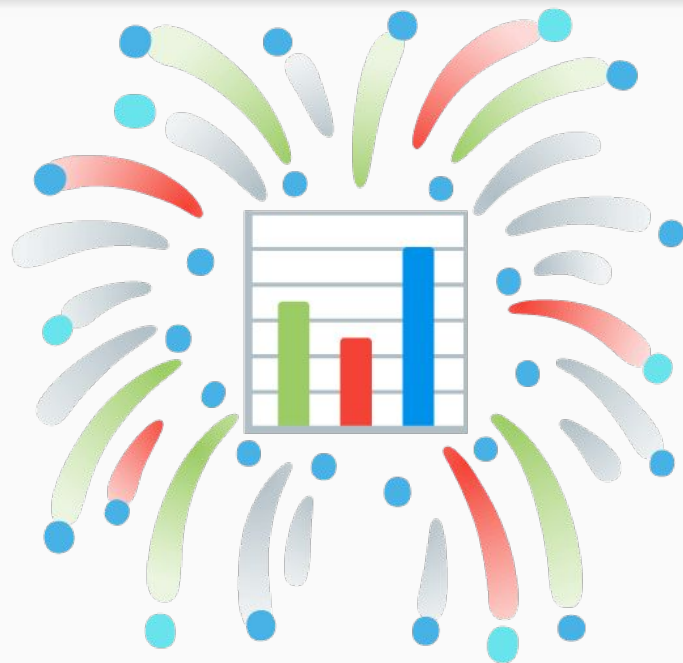
# Setting Up

- Log in, request a server
  - Either CPU or GPU instances



# Setting Up

- Log in, request a server
  - Either CPU or GPU instances
- Create a new mamba env
- Install nugraph dependencies
- Run away!



# Dependency Wrangling

- Using mamba is recommended

- On first time running:

```
$ mamba init  
$ echo "if [ -f ~/.bashrc ]; then . ~/.bashrc fi" > .bash_profile
```

- Install dependencies as normal

```
$ git clone https://github.com/nugraph/nugraph.git  
$ cd nugraph  
$ mamba install -c pyg pyg  
$ mamba install -c numl pynuml  
$ mamba install conda-forge nvidia pytorch  
$ pip install --no-deps -e ./nugraph/
```

- Additionally, install `ipykernel`

- Needed to load the environment into a notebook (if that's your jam)



# Data and checkpoints

- NuGraph has its own Ceph filesystem
  - /exp/projects/fwk/nugraph/
  - Contains nugraph2 and 3 data, checkpoints from nugraph2
- Burt is very nice and gave us a faster access disk for nugraph2 data
  - /nugraph/

# Things of Note

- EAF has no ssh access
  - If you need to connect to a server, it has to be done through a web protocol
- Your code will need to be present on the remote machine to be run
- If you disconnect from the EAF while running your code in a terminal, your terminal session ends



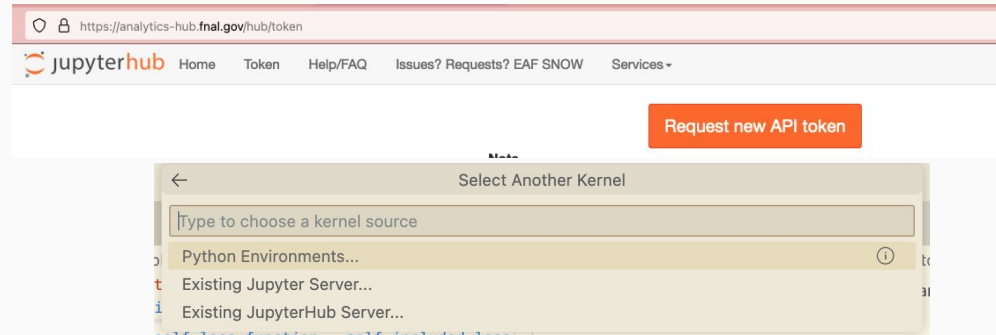
# Training in Notebooks

- On the web EAF

- Clone your code from a remote repository (recommended), or upload and unzip
- Either execute scripts using line magic (start a line with “!”) or write bespoke python code

- Using vscode

- Request a token
- In vscode, select a kernel and select ‘Existing JupyterHub Server’
- Enter



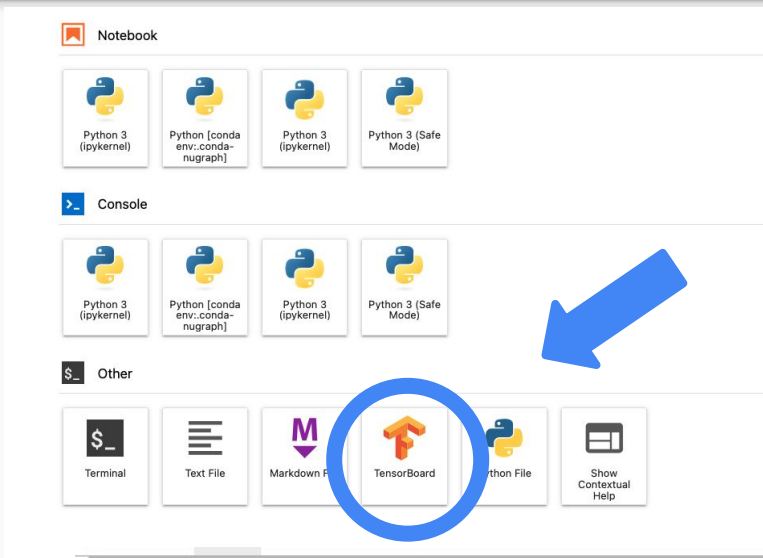
`https://analytics-hub.fnal.gov/user/{your-un}/?token={Your new token}`

# Training in Scripts

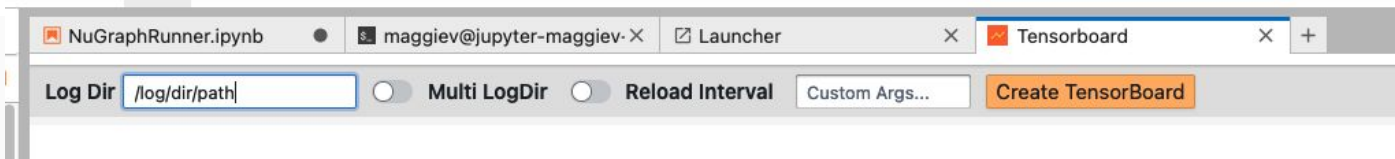
1. Just run things normally. There's nothing special here. It's an on-demand computing resource.
2. Open a terminal window and go for it.
3. Submit to fermigrid using `jobsub_submit` once you have a kerberos token



# Using Tensorboard



- Tensorboard runs natively on most instances
- Logs can be anywhere, but best to keep them in your home directory.

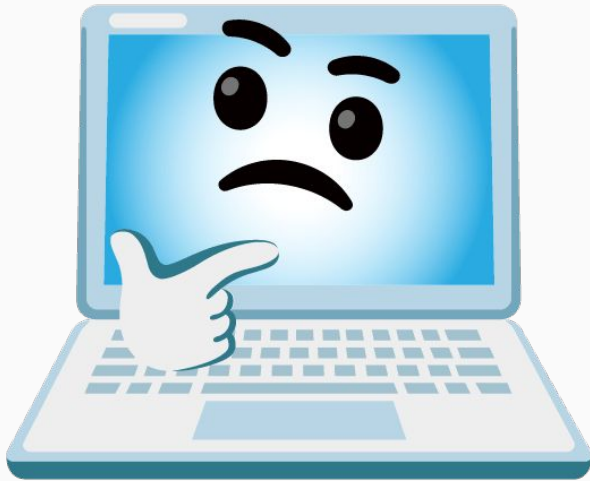


# Extra Info

- Sessions time out after 24 hours of inactivity
  - Running in a script means you need to have a notebook running something (even if it's just printing blank lines or sleeping)
- Your home directory has limited space, and if you run out, you will be unable to launch a server
  - Launch a rescue image and start deletin'



# Documentation and Resources



- <https://eafjupyter.readthedocs.io/>
- Join #eaf-users in [fnal.slack.com](https://fnal.slack.com)
- Join the email list - [eaf-users@fnal.gov](mailto:eaf-users@fnal.gov)