Fermilab U.S. DEPARTMENT OF ENERGY Office of Science

AI infrastructure: the CSAID view

Ken Herner, CSAID, Scientific Computing Systems and Services Division

AI Infrastructure planning workshop

6 Apr 2023

So you need a GPU…

- AI Practitioners have ready* access to the following resources
  - Elastic Analysis Facility
  - Wilson Cluster, through experiments and/or dedicated projects
  - GPU clusters at remote sites via usual FIFE job submission tools through experiments
  - HPC centers such as NERSC through experiments and HEPCloud
  - Other HPC centers with individual/university/experiment allocations (CAN be integrated with HEPCloud!)
- Each type has slightly different strengths and weaknesses, along with different interfaces
- We will (briefly!) discuss each in turn, and where to go for more information

"When you come to a fork in the road, take it." - L. P. "Yogi" Berra

*commercial clouds are also possible, though access is less ready (and no large common pool available)

Elastic Analysis Facility

- Jupyter Hub deployment with general CPU and GPU-enabled notebooks available. Highly scalable, customizable, and replicable elsewhere
- GPUs are available through the login interface (must be on-site or on VPN for now)
  - They are finite of course, but A100s now deployed
- Latest documentation is here. In particular, look at the server and notebook options section to see what's in each flavor (CVMFS also available)

EAF (2)

- How do I get an account?
  - Anyone with a services account can log in, but follow your experiment's usual instructions
- Can I customize the environment and packages?
  - Yes. Several things are possible with preamble scripts, mamba, pip, etc. See the instructions about customizing
- What storage is available?
  - 24 GiB work area in Ceph as home area, usual /nashome areas (FIFE/DUNE/Cosmic), usual LPC /uscms NFS areas also available
  - Working on mounting neutrino/muon NAS areas
  - Streaming with xrootd also works to access larger storage elements

GPUs via Wilson Cluster

- The Wilson Cluster is available to all of FNAL either through experiments/departments or specific projects (focused on a particular area; could be sub-project of an experiment, but do *not* have to be part of something larger)
- Available GPUs range from K80s (100) to A100s (4); base OS is SL7
- More information available at https://computing.fnal.gov/wilsoncluster/
  - There is a SNOW form to be added to a project (automatically added to base experiment projects)
- Rumors persist that queue times are interminable for the GPUs: not true as of today

SLURM Partition	SLURM GPU Resource:Type:Numberof	Processor	Nodes	Total Threads	Memory/Core	Memory/Host	GPU
cpu_gce	No GPU	2.6GHz Dual CPU Eight Core Intel	108	16	8GB	128GB	None
cpu_gce_test	No GPU	2.6GHz Dual CPU Eight Core Intel	7	16	8GB	128GB	None
gpu_gce	gpu:k40:4	2.6GHz Dual CPU Eight Core Intel	4	16	8GB	128GB	4x NVIDIA Kepler K40
	gpu:p100nvlink:2	2.4GHz Dual CPU Fourteen Core Intel	1	56	2GB	112GB	2x NVIDIA Pascal P100 (w/NVLINK)
	gpu:p100:8	1.7GHz Dual CPU Eight Core Intel	1	16	48GB	768GB	8x NVIDIA P100
	gpu:v100:2	2.5GHz Dual CPU Twenty Core Intel	4	40	4.7GB	188GB	2x NVIDIA Volta V100
	gpu:a100:4	2.8GHz Dual CPU Thirty-Two Core EPYC 7543	1	64	8GB	512GB	4x NVIDIA Ampere A100-80 80GB HBM2e

cpu_gce | cpu_gce_test | gpu_gce | gpu_gce_ppc | gpu_ose | knl_gce | Project Usage

Refresh Rate=5 mins. Last updated on: 04-05-2023 1:05 AM

SLURM job status for partition gpu_gce

JobID	Job Name	User	Account	# of Nodes	Partition	QoS	List of Nodes -or- (Reason for job queued)	Priority	GPU Resources	Start Time	Requested Walltime (DAY-HH:MM:SS)	Time Remaining (DAY-HH:MM:SS)	Job State
408879	ecd_gpu	xinyuan	fwk	1	gpu_gce	regular	wcgpu04	67492	gres:gpu:1	2023-04-04T06:10:52	1-00:00:00	5:05:47	RUNNING
408880	ecd_gpu	xinyuan	fwk	1	gpu_gce	regular	wcgpu04	67491	gres:gpu:1	2023-04-04T06:10:52	1-00:00:00	5:05:47	RUNNING
408958	bash	Soon yung Jun	g4p	1	gpu_gce	opp	wcgpu05	339	gres:gpu:1	2023-04-04T17:06:55	8:00:00	1:50	RUNNING
407393	diffu_dataset3_mar21_hybrid_n400	oamram	cms	1	gpu_gce	opp	wcgpu03	44752	gres:gpu:1	2023-04-04T17:35:20	8:00:00	30:15	RUNNING
408966	train	calcuttj	dune	1	gpu_gce	opp	wcgpu06	19260	gres:gpu:v100:2	2023-04-05T00:17:21	8:00:00	7:12:16	RUNNING
407392	diffu_dataset3_mar21_hybrid_weight_n400	oamram	cms	1	gpu_gce	opp	wcgpu05	46830	gres:gpu:1	2023-04-05T00:51:22	8:00:00	7:46:17	RUNNING

Fermilab

GPUs on HTC Clusters

- Several sites have GPUs available; accessible via usual job submission commands for IF/CF expts (separate for CMS)
  - Simply add --lines='+RequestGPUs=1' to your submission.
  - **As of now, user is responsible for specifying appropriate container** (use the --singularity-image option and make sure the image is available to the worker, usually via CVMFS). One could imagine the division maintaining a container including CUDA, IF there is demand for it. Must also be sure data transfer in/out is possible
- Hardware ranges from a few years old up to A100s depending on the site. Additional HTCondor classads are possible to match if you need a specific type of GPU. Contact FIFE for details.
- Several advantages to using these resources
  - Very little competition for them (i.e. no long queues a la NERSC)
  - No finite allocations!
- GPGrid has no GPU nodes now; could consider if demand (more later)

NERSC and HEPCloud

- Perlmutter (both CPU and GPU queues) available via HEPCloud for currently onboarded expts (CMS, DUNE, g-2, Microboone, Mu2e, Nova)
  - Contact your experiment's production group
- Demonstrations of Inference server setups in progress
- No true "co-scheduling" of GPU servers w/ other workflows on non-GPU queues yet
  - Ad-hoc or requires partitioning between workflow components
- No support for MPI workflows currently within HEPCloud
  - Need to rely on local slurm batch queue

NERSC and HEPCloud (2), other HPC centers

- It is also possible to integrate other HPC center or commercial cloud allocations into HEPCloud; contact HEPCloud team
  - Advantages include not having to learn separate infrastructures at every different site
- Some HPC centers offer testbed platforms for AI/ML apps; see ANL as one example

Future Hardware Additions

- Lattice QCD is working on a new cluster, LQ2, which will be GPU-based. Expected later this year (PO went out this week)
- Aside from that, CSAID is planning to buy roughly $500k of GPU workers equipped with A100s
  - Form factor not yet decided (2 or 4 GPUs per node)
  - Expect this will provide up to a few dozen GPUs
  - Looked at Hopper cards but decided better bang for the buck with A100s given our typical models
- Machines will have NVLink, Infiniband, ethernet
- Distribution among the various clusters not decided yet. Depends a bit on needs, demand, and stakeholder requirements
  - *One could imagine* shifting machines between clusters as demand ebbs and flows between different areas (assuming sufficient lead time)

Summary

- Wide variety of resources now available from interactive containers to grid slots to large blocks on HPC centers
- Different tools for different jobs: don’t hesitate to contact division personnel even if it’s just to consult about what’s the best option
- Planning a large purchase later this year and we need to figure out how to deploy it:

WE NEED TO HEAR FROM YOU!!!