OAK RIDGE National Laboratory

NOAA GPU Hackathon – Ascent Overview

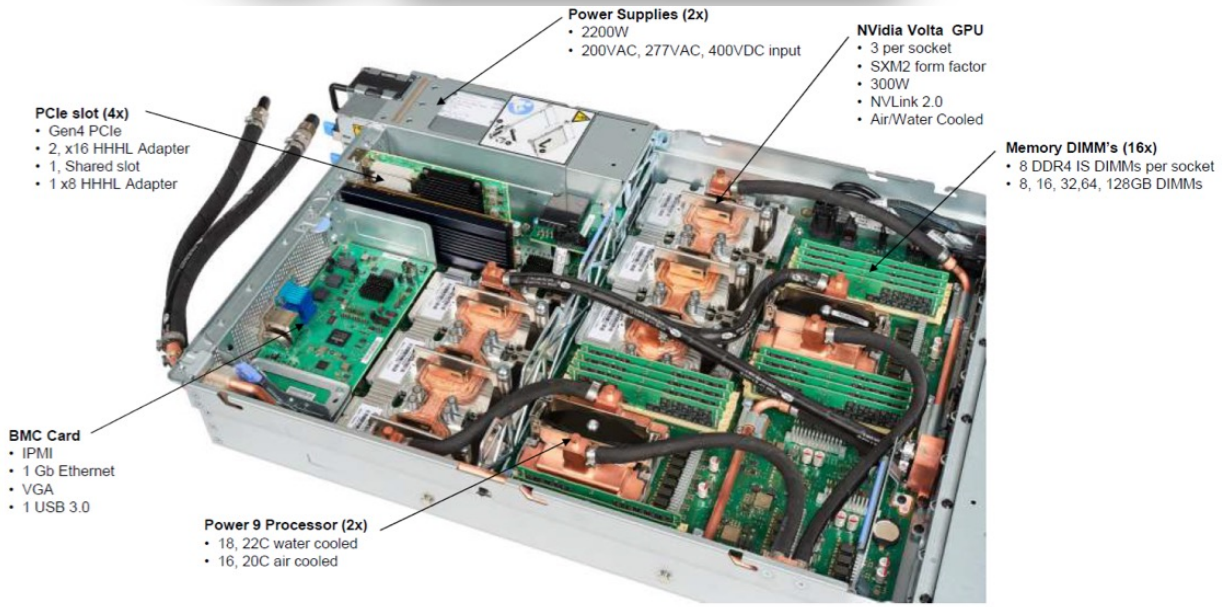
Brian Smith

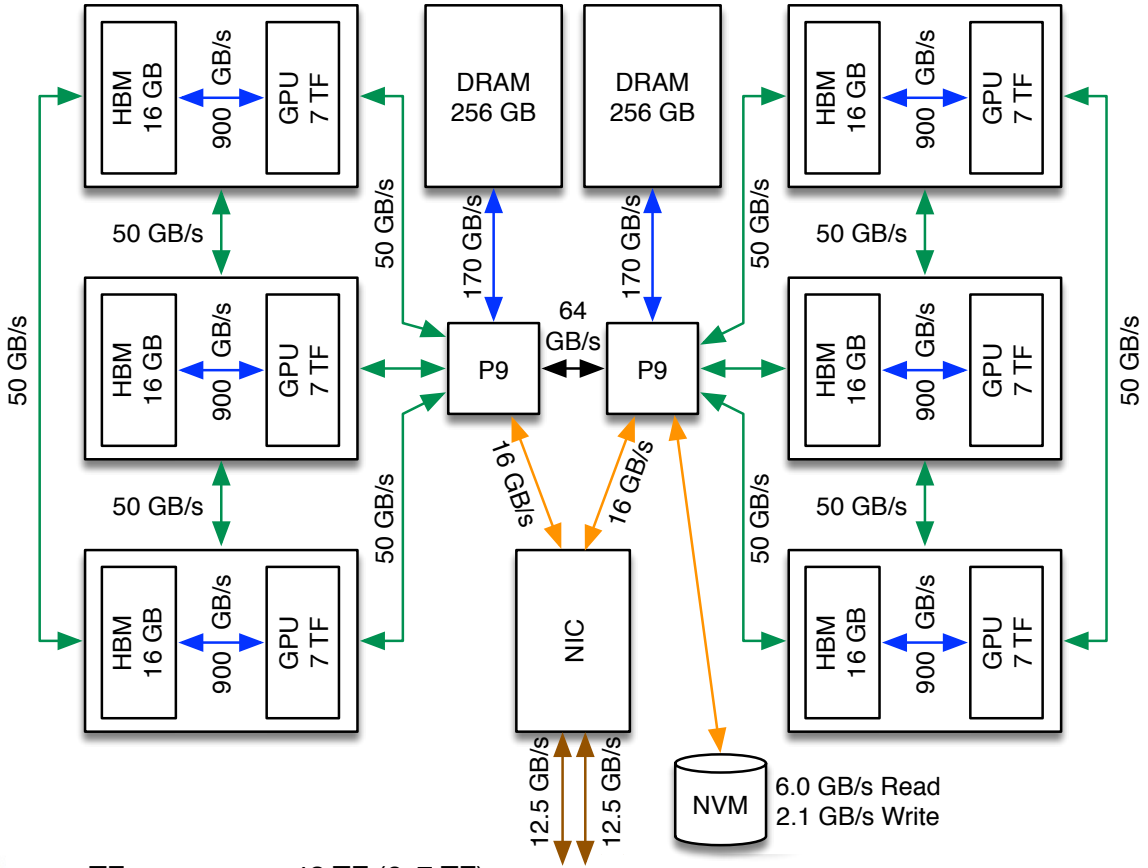
Oak Ridge Leadership Computing Facility

August 12, 2021

U.S. DEPARTMENT OF ENERGY

summit





TF	42 TF (6x7 TF)
HBM	96 GB (6x16 GB)
DRAM	512 GB (2x16x16 GB)
NET	25 GB/s (2x12.5 GB/s)
MMsg/s	83

- HBM/DRAM Bus (aggregate B/W)
- NVLINK
- X-Bus (SMP)
- PCIe Gen4
- EDR IB

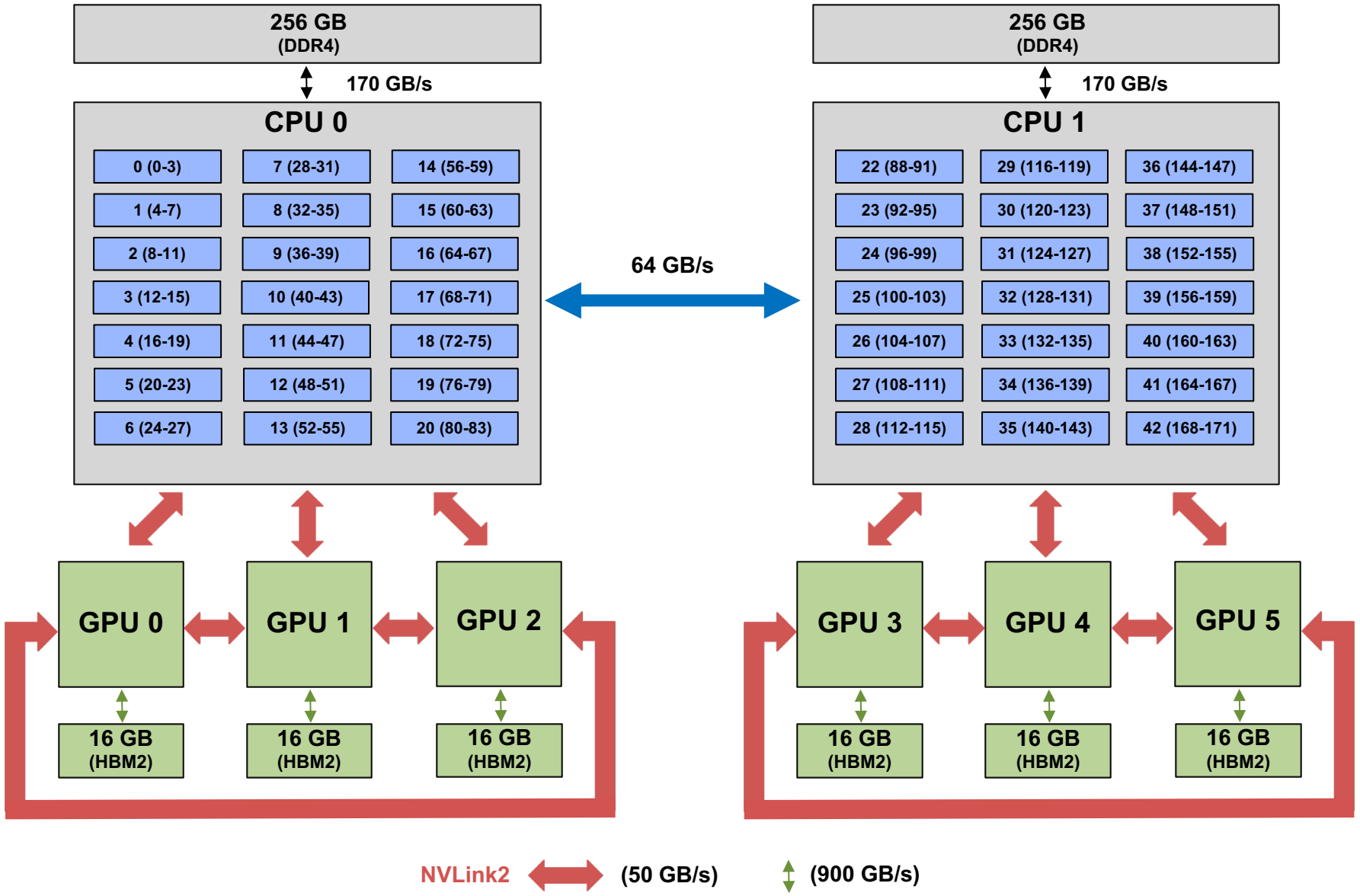
HBM & DRAM speeds are aggregate (Read+Write).
All other speeds (X-Bus, NVLink, PCIe, IB) are bi-directional.

Node Overview

OAK RIDGE National Laboratory

Summit Node

(2) IBM Power9 + (6) NVIDIA Volta V100



Available File Systems / Storage Areas on Ascent

NFS Directories – This is where you might want to keep source code and build your application.

NOTE: These directories are read-only from the compute nodes!

`/ccsopen/home/userid`

- Your personal home directory

`/ccsopen/proj/gen040`

- Can be accessed by all participants of this event
- You should create a directory here with your team name to collaborate (source code, scripts, etc.)

GPFS Directories (parallel file system) – This is where you should write data when running on Ascent's compute nodes.

`/gpfs/wolf/gen040/scratch/userid`

- Your personal GPFS scratch directory

`/gpfs/wolf/gen040/proj-shared`

- Can be accessed by all participants of the event
- You should create a directory here with your team name to collaborate (data written from compute nodes)

jsrun – Basic Options

```
jsrun [-n #resource sets] [CPU cores, GPUs, tasks in each resource set] program [program args]
```

jsrun Flags		Description	Default Value
Long	Short		
--nrs	-n	Number of RS	All available physical cores
--tasks_per_rs	-a	Number of MPI tasks (ranks) per RS	N/A (total set instead [-p])
--cpu_per_rs	-c	Number of CPUs (physical cores) per RS	1
--gpu_per_rs	-g	Number of GPUs per RS	0
--bind	-b	Number of physical cores allocated per task	packed:1
--rs_per_host	-r	Number of RS per host (node)	N/A
--latency_priority	-l	Controls layout priorities	gpu-cpu,cpu-mem,cpu-cpu
--launch_distribution	-d	Order of tasks started on multiple RS	packed

See `man jsrun` for full list of options

`jsrun` Job Launcher – Tools & Documentation

`hello_jsrun`

- https://code.ornl.gov/t4p/Hello_jsrun
- Simple "Hello World"-type program used to test layout of resources on a Summit node using `jsrun`.
- As stated in the README, be sure to start MPS server (`-alloc_flags "gpumps"`)

`job-step-viewer`

- https://jobstepviewer.olcf.ornl.gov/

"Job Launcher (jsrun)" section of the Summit User Guide

- https://docs.olcf.ornl.gov/systems/summit_user_guide.html#job-launcher-jsrun

Additional training materials to learn about `jsrun`

A (fairly) quick tutorial on using the `jsrun` job launcher on Summit/Ascent:

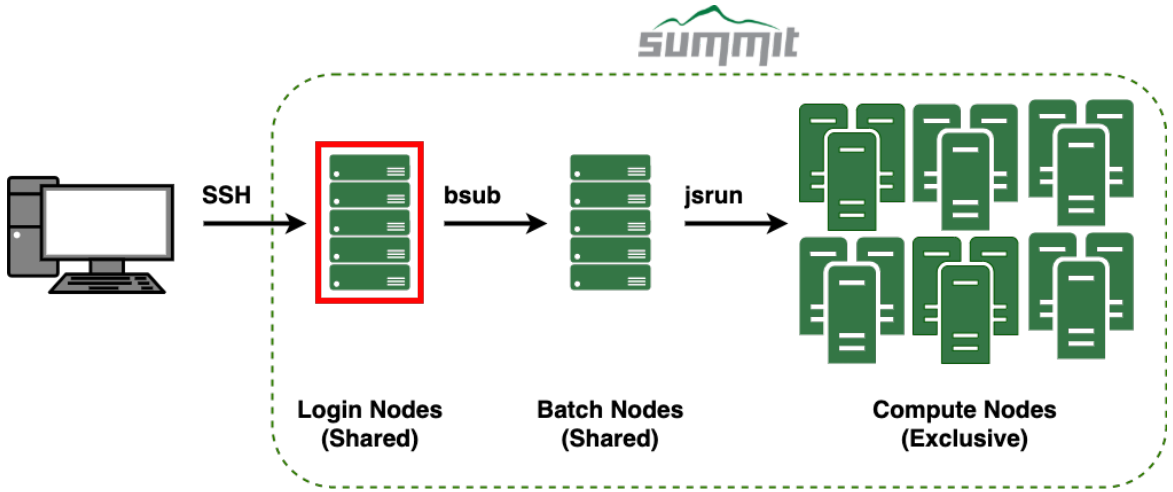
- https://github.com/olcf-tutorials/jsrun_quick_start_guide

Recent Presentation of "jsrun Basics"

- Slides: https://www.olcf.ornl.gov/wp-content/uploads/2019/12/jsrun_basics.pdf
- Recording: https://vimeo.com/393782415

Must use `jsrun` to run on compute nodes

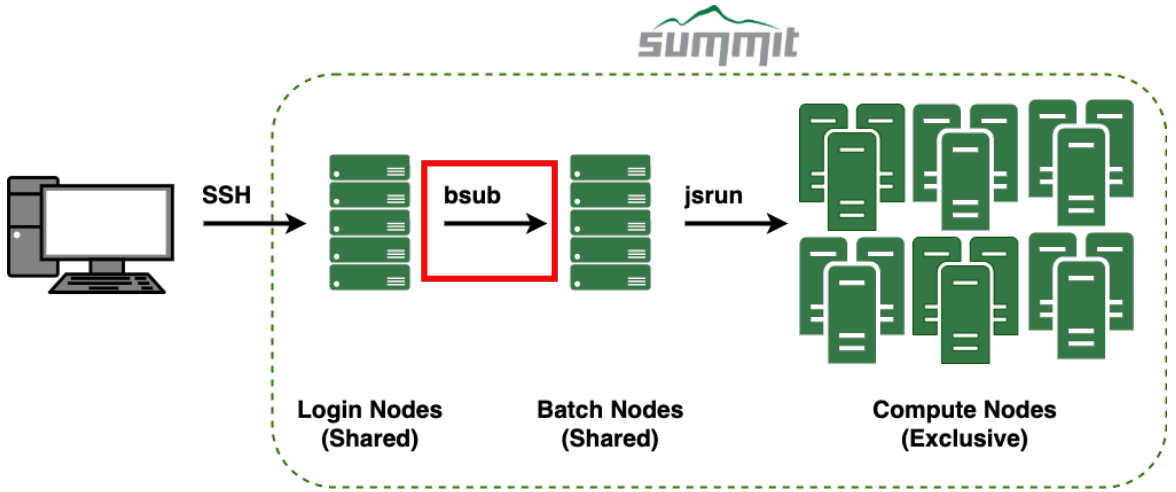
```
[t4p@login1: ~]$ hostname
login1
```



Must use `jsrun` to run on compute nodes

```
[t4p@login1: ~]$ hostname
login1

[t4p@login1: ~]$ bsub -P GEN040 -nnodes 1 -W 60 -Is /bin/bash
Job <15167> is submitted to default queue <batch>.
<<Waiting for dispatch ...>>
<<Starting on login1>>
```

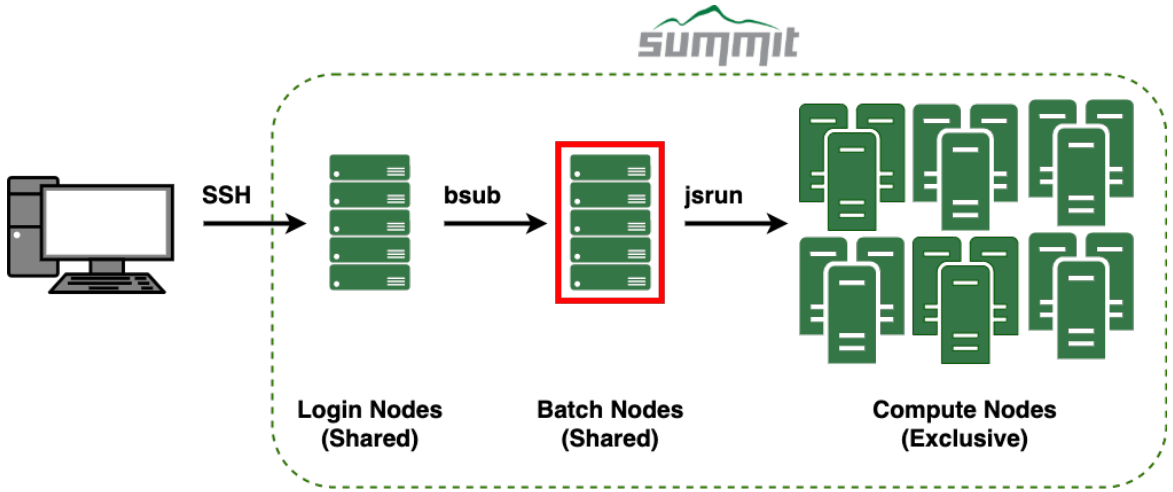


Must use `jsrun` to run on compute nodes

```
[t4p@login1: ~]$ hostname
login1

[t4p@login1: ~]$ bsub -P GEN040 -nnodes 1 -W 60 -Is /bin/bash
Job <15167> is submitted to default queue <batch>.
<<Waiting for dispatch ...>>
<<Starting on login1>>

[t4p@login1: ~]$ hostname
login1
```



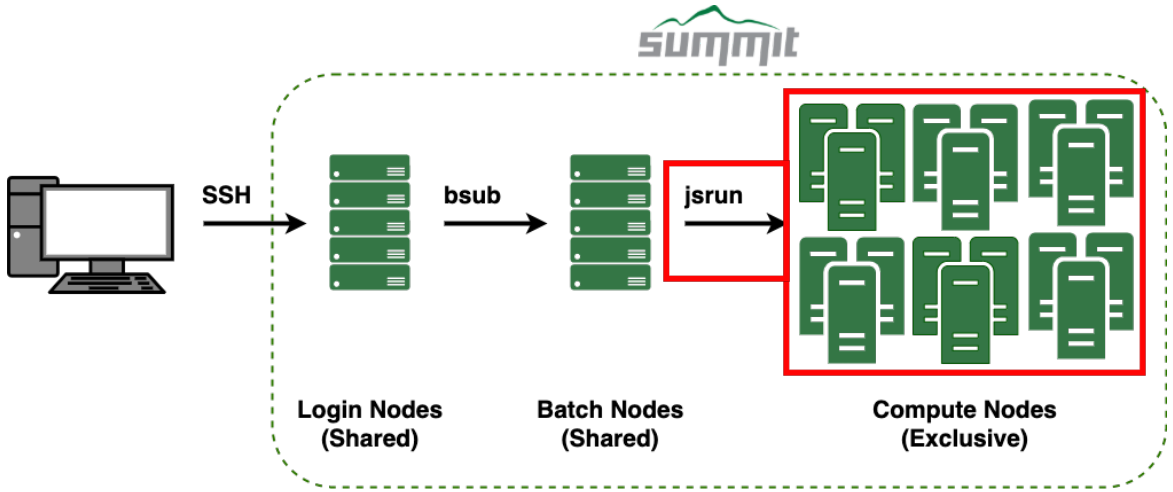
Must use `jsrun` to run on compute nodes

```
[t4p@login1: ~]$ hostname
login1

[t4p@login1: ~]$ bsub -P GEN040 -nnodes 1 -W 60 -Is /bin/bash
Job <15167> is submitted to default queue <batch>.
<<Waiting for dispatch ...>>
<<Starting on login1>>

[t4p@login1: ~]$ hostname
login1

[t4p@login1: ~]$ jsrun -n1 hostname
h49n16
```



The login nodes are shared among all participants (compiling, file editing, data analysis, etc.), so please **DO NOT RUN YOUR APPLICATIONS ON THE LOGIN NODES!!**

Some Useful Commands / Flags

`jobstat`

- Shows information about the jobs running on the system

`--smpiargs="-gpu"`

- `jsrun` flag that enables CUDA-Aware MPI
- If you are not familiar with CUDA-Aware MPI or GPUDirect, please see this tutorial: https://github.com/olcf-tutorials/MPI_ping_pong

`-alloc_flag "gpumps smt1"`

- `bsub` flag that allows you to start a CUDA MPS server or change the SMT mode of the physical CPU cores
- Multiple options are separated by a space-delimited list

Ascent Queue Policy

Number of Nodes	Max Walltime
1 – 2	2 hours
3 – 4	1 hour

There are a total of 16 schedulable compute nodes in Ascent, so please be respectful of others when requesting resources…

- Try to limit yourself to 1 compute node unless needed
- When you're finished with an allocation, please kill it (i.e., `exit` from within an interactive job or `bkill JOBID` for batch jobs).

Other Helpful Links

OLCF Summit User Guide

- https://docs.olcf.ornl.gov/systems/summit_user_guide.html
- NOTE: Ascent mounts different file systems than Summit, so please refer to info in these slides or the Training System (Ascent) section of the Summit User Guide for this information
  - https://docs.olcf.ornl.gov/systems/summit_user_guide.html#training-system-ascent
- NVIDIA's Nsight Profiling Tools
  - https://docs.olcf.ornl.gov/systems/summit_user_guide.html#profiling-gpu-code-with-nvidia-developer-tools

OLCF Training Archive

- Contains slides and recordings from previous OLCF training events.
- https://docs.olcf.ornl.gov/training/training_archive.html

Accounts

- Accounts will be active until Friday, Sept 10 2021

ML/AI/DL On Ascent

- IBM Provides the Open-CE Python Conda environment
  - Contains many common AI tools, e.g. TF, pyTorch, etc
  - More details here: https://docs.olcf.ornl.gov/software/analytics/ibm-wml-ce.html

Questions?