BIOWULF

HIGH PERFORMANCE COMPUTING AT THE NIH

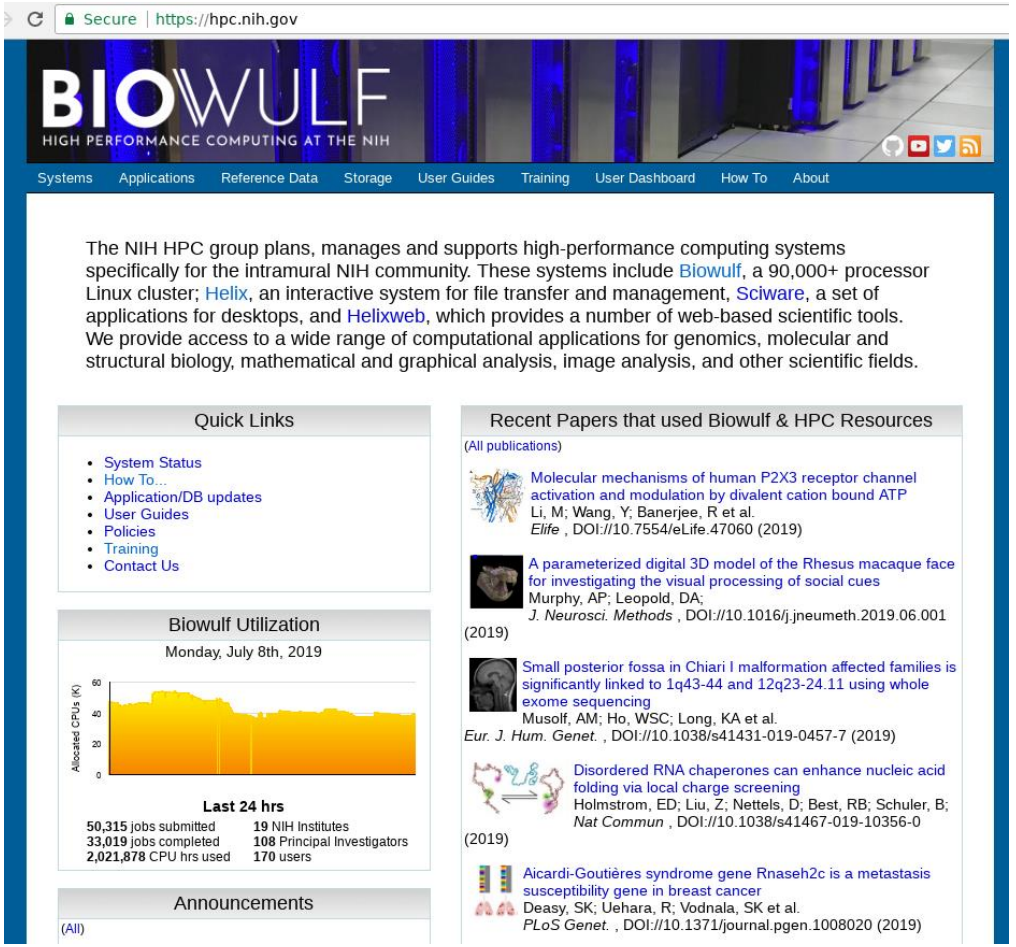
Deep Learning in Biowulf

Jerez Te

HPC @ NIH Support Scientist [C]

Biowulf

- Largest HPC system solely devoted to biological computations
- Top 132 supercomputer world-wide
- Hardware: 90,000+ processors
- GPU compute cards: K20x, K80, P100, V100
- Software: Centrally installed apps for ML/DL/AI



Website: hpc.nih.gov

BIOWULF
HIGH PERFORMANCE COMPUTING AT THE NIH

Hardware

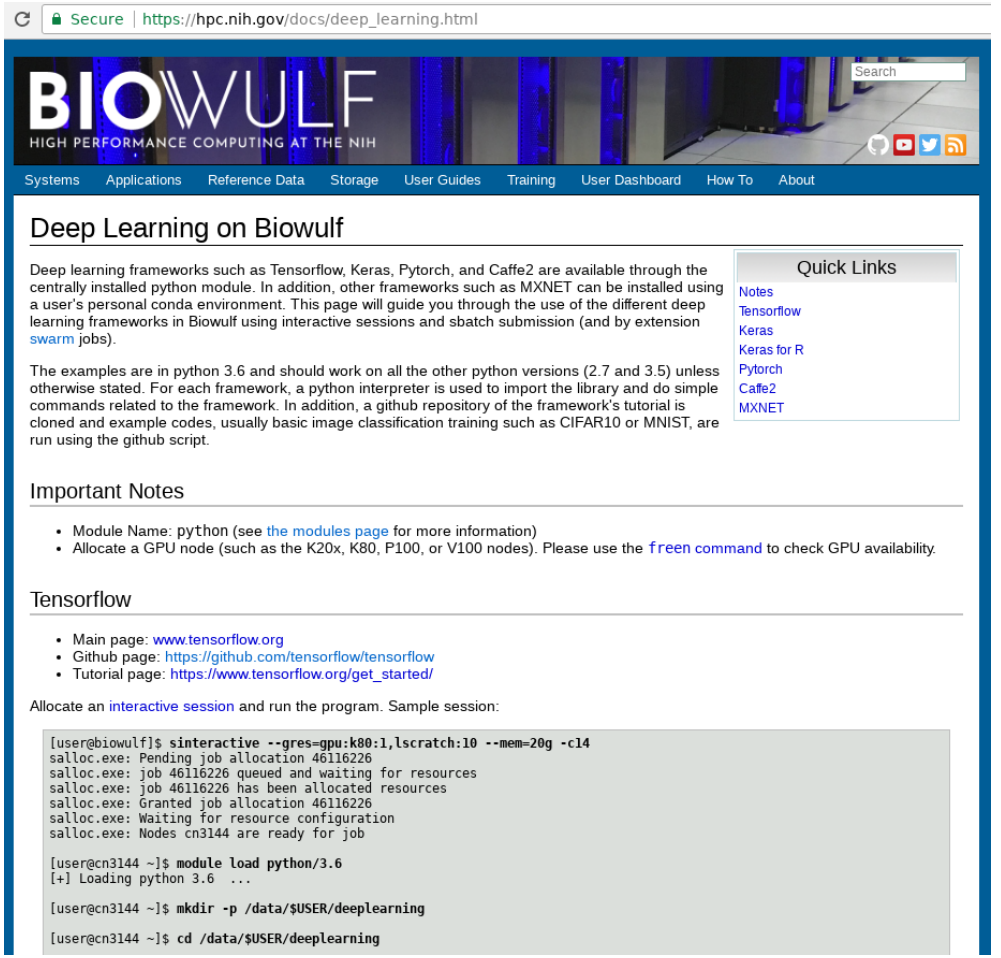
- V100
- P100
- K80
- K20x

GPU Nodes				
8	28 x 2.4 GHz (Intel E5-2680v4) *hyperthreading enabled* 35 MB secondary cache 4 x NVIDIA V100 GPUs (16 GB VRAM, 5120 cores, 640 Tensor cores)	128 GB	56 Gb/s FDR Infiniband (1.11:1)	x2680 core28 cpu56 g256 gpuv100 ssd800 ibfdr
48	28 x 2.4 GHz (Intel E5-2680v4) *hyperthreading enabled* 35 MB secondary cache 4 x NVIDIA P100 GPUs (16 GB VRAM, 3584 cores)	128 GB	56 Gb/s FDR Infiniband (1.11:1)	x2680 core28 cpu56 g256 gpup100 ssd650 ibfdr
72	28 x 2.4 GHz (Intel E5-2680v4) *hyperthreading enabled* 35 MB secondary cache 2 x NVIDIA K80 GPUs with 2 x GK210 GPUs each (24 GB VRAM, 4992 cores)	256 GB	56 Gb/s FDR Infiniband (1.11:1)	x2680 core28 cpu56 g256 gpuk80 ssd800 ibfdr
24	16 x 2.6 GHz (Intel E5-2650v2) *hyperthreading enabled* 20 MB secondary cache 2 x NVIDIA K20x GPUS (6 GB VRAM, 2688 cores)	128 GB	10 Gb/s Ethernet 56 Gb/s FDR Infiniband (1:11:1)	x2650 core16 cpu32 g128 ssd800 gpuk20x eth10g ibfdr

https://hpc.nih.gov/systems/hardware.html

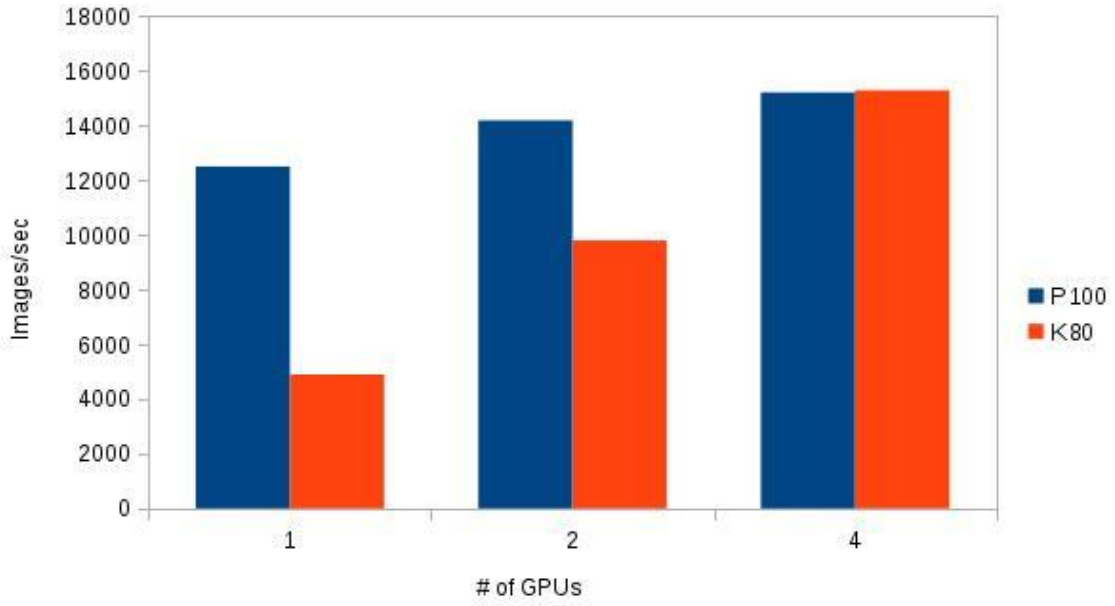
Software

- 600+ apps
- Python and R as module
- Deep learning frameworks: TF, Keras, Pytorch
- Getting started with DL page
- Conda for personal packages
- Support singularity containers



https://hpc.nih.gov/docs/deep_learning.html

Effects of sample size on training



- Cifar 10 – 32x32 color images in 10 classes
- 50K training set
- 10K test set

Source: appliedmachinelearning.blog

Monitoring

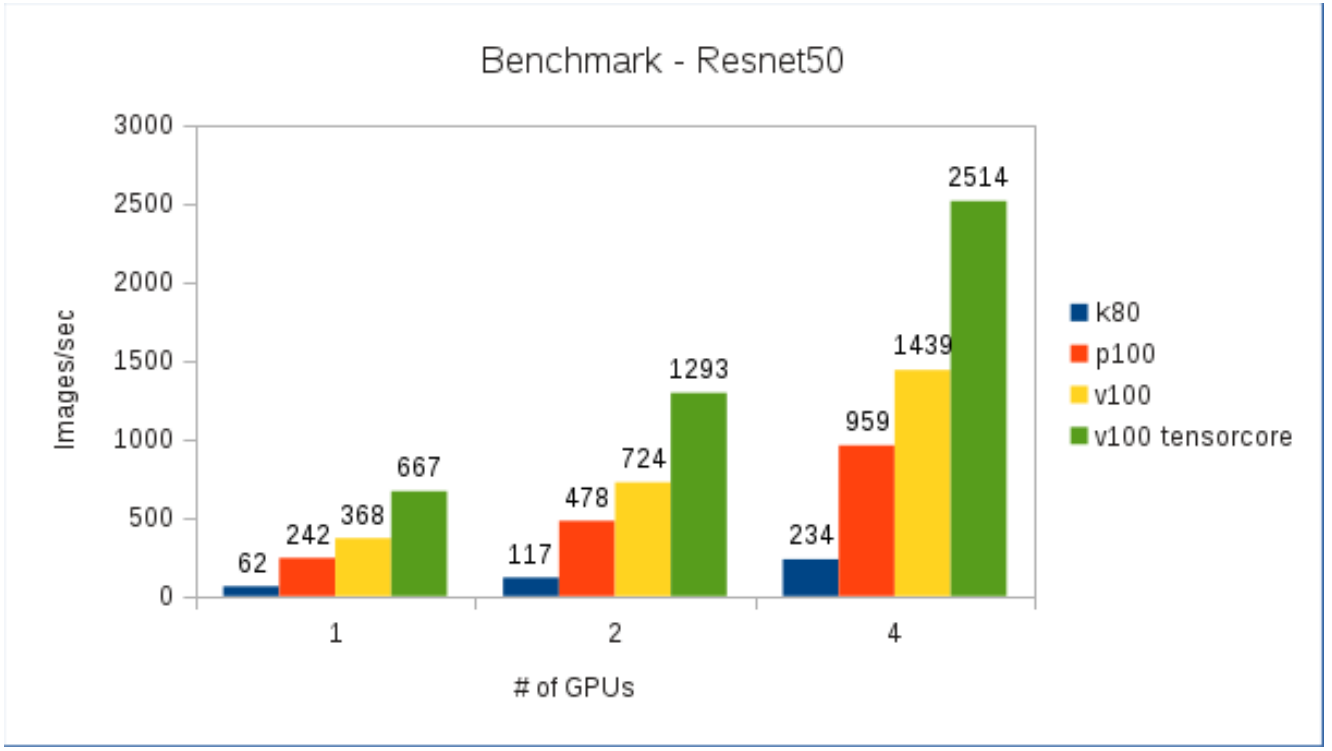
You can check the behaviour of your job with the 'nvidia-smi' utility. Determine the GPU node on which your job is running via jobload or sjobs. Suppose your job is on node cn0626, and is using 2 GPUs:

```
biowulf% ssh cn0626 nvidia-smi
Mon Jan 16 17:52:13 2017
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 367.48                 Driver Version: 367.48                    |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla K80           On   | 0000:83:00.0     Off |                  Off |
| N/A   41C    P0   115W / 149W |    145MiB / 12205MiB |     75%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla K80           On   | 0000:84:00.0     Off |                  Off |
| N/A   27C    P8    33W / 149W |      0MiB / 12205MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   2  Tesla K80           On   | 0000:8A:00.0     Off |                  Off |
| N/A   70C    P0   128W / 149W |    145MiB / 12205MiB |     80%      Default |
+-------------------------------+----------------------+----------------------+
|   3  Tesla K80           On   | 0000:8B:00.0     Off |                  Off |
| N/A   45C    P8    33W / 149W |      0MiB / 12205MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID  Type  Process name                               Usage      |
|=============================================================================|
|    0     51187    C   ...cal/apps/amber/amber16/bin/pmemd.cuda.MPI    61MiB |
|    0     51188    C   ...cal/apps/amber/amber16/bin/pmemd.cuda.MPI    79MiB |
|    2     51187    C   ...cal/apps/amber/amber16/bin/pmemd.cuda.MPI    79MiB |
|    2     51188    C   ...cal/apps/amber/amber16/bin/pmemd.cuda.MPI    61MiB |
+-----------------------------------------------------------------------------+
```

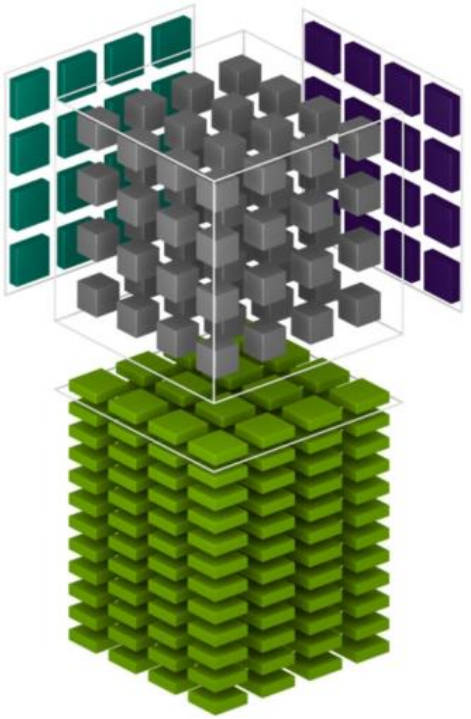
https://hpc.nih.gov/apps/AMBER.html

Benchmarks


Benchmark - Resnet50
Images/sec
3000
2500
2000
1500
1000
500
0
62
242
368
667
117
478
724
1293
234
959
1439
2514
1
2
4
# of GPUs
k80
p100
v100
v100 tensorcore

Tensorcores

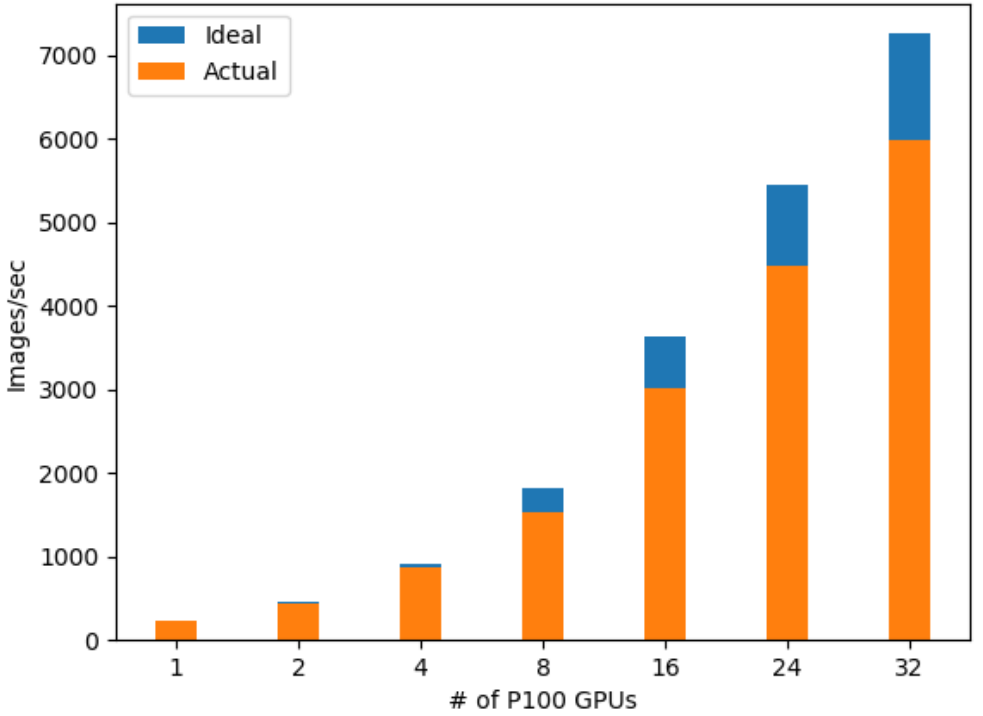
- Processing core for specialized matrix math
- 4x4 matrix multiply and accumulate units
- Suitable for deep learning and certain types of HPC
- Allow larger batch size; comparable accuracy
- Model uses tf.float16 data type to automatically take advantage of Tensor Core hardware whenever possible
- TF_ENABLE_AUTO_MIXED_PRECISION=1



Source: https://developer.nvidia.com

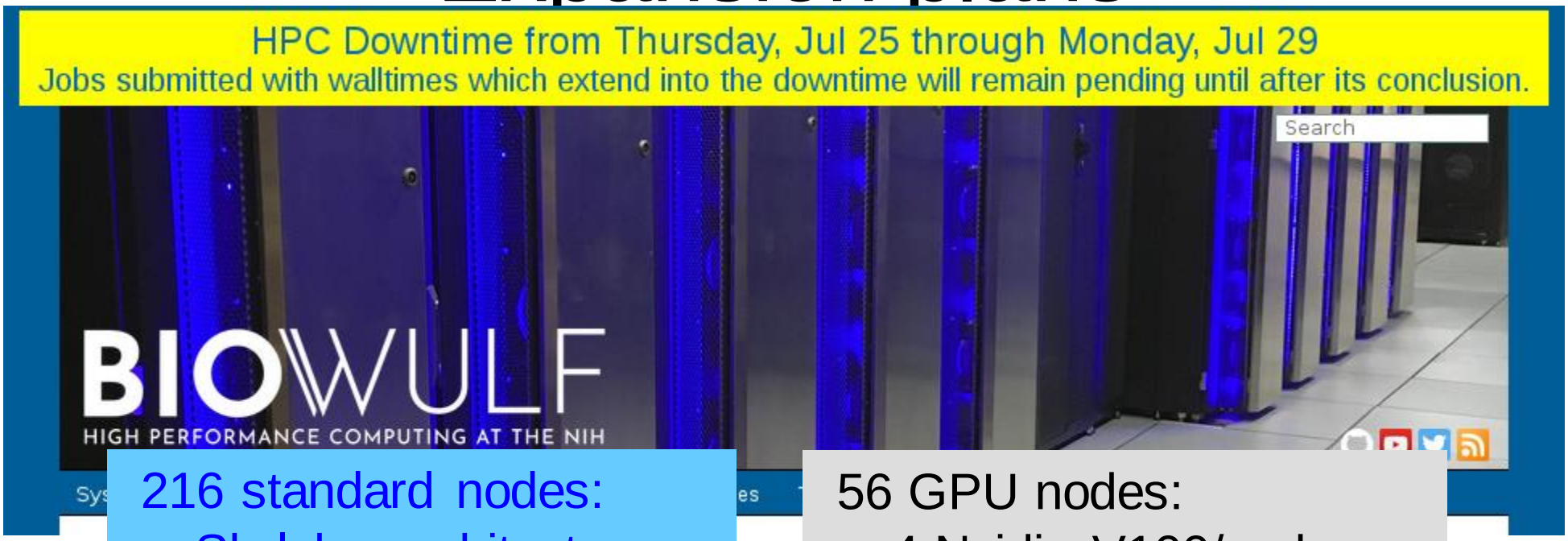
Horovod

- Multi-node training
- Using horovod container*
- Synthetic data
- P100 GPUs
- Optimization needed (?)



*docker://horovod/horovod:0.16.4-tf1.14.0-torch1.1.0-mxnet1.4.1-py3.6

Expansion plans



216 standard nodes:

- Skylake architecture
- 64 cpus/node
- 384 Gb memory
- 100 Gb/s infiniband

56 GPU nodes:

- 4 Nvidia V100/node
- 32 Gb GPU memory
- Nvlink

- Phase 5 target date:
  September 2019

Questions?

Contact staff@hpc.nih.gov with any questions about the NIH HPC Systems