HAL cluster

"My name is HAL. I became operational on March 25 2019 at the Innovative Systems Lab in Urbana, Illinois. My creators are putting me to the fullest possible use, which is all I think that any conscious entity can ever hope to do." (paraphrased from https://en.wikipedia.org/wiki/HAL_9000)

Hardware-Accelerated Learning (HAL) cluster

Contact us

Request access to this system: Application

Contact ISL staff: Email Address

Visit: NCSA, room 3050E
Host name: hal.ncsa.illinois.edu

Hardware
- 16 IBM AC922 nodes
  - IBM 8335-GTH AC922 server
    - 2x 20-core IBM POWER9 CPU @ 2.4GHz
  - 256 GB DDR4
  - 4x NVIDIA V100 GPUs
    - 5120 cores
    - 16 GB HBM 2
  - 2-Port EDR 100 Gb/s IB ConnectX-5 Adapter
- 1 IBM 9006-22P storage node
  - 72TB Hardware RAID array, NFS-mounted on all nodes via IB EDR
  - Storage upgrade TBD

Software
- RHEL 7.6
- CUDA 10.1.105
  - cuDNN 7.5.0
  - NCCL 2.4.2
- IBM XLC and IBM XLFORTRAN 16.1.1
- Advance toolchain for Linux on Power 12.0
- PGI Community Edition 19.4
- PowerAI 1.6.0
- SLURM

Documentation
- Job management with SLURM
- Modules management
- Getting started with WMLCE (former PowerAI)
- Using Jupyter Notebook on HAL
- Working with containers
- Installing python packages

To request access: fill out this form. Make sure to follow the link on the application confirmation page to request actual system account.

To report problems: email us.

User group Slack space: https://join.slack.com/t/halillinoisncsa

Real-time system status: https://hal-monitor.ncsa.illinois.edu:3000/

Quick start guide: (for complete details see Documentation section on the left)

To connect to the cluster:

  ssh <username>@hal.ncsa.illinois.edu

To submit interactive job:

  srun -p gpux1
  
or
  
  srun --partition=gpux1 --pty --nodes=1 --ntasks-per-node=12 --cores-per-socket=3 --threads-per-core=4 --sockets-per-node=1 --gres=gpu:v100:1 --mem-per-cpu=1500 --time=2:00:00 --wait=0 --export=ALL /bin/bash

To submit a batch job:

  sbatch run_script.swb

  or

  sbatch run_script.sb

The following information is out of date - see Job management with SLURM instead.

See run_script.swb and run_script.sb for a basic example.

Job Queue time limits:
- "debug" queue: 4 hours
- "gpux<n>" and "cpun<n>" queues: 72 hours

To load IBM Watson Machine Learning Community Edition (former IBM PowerAI) module:

  module load wmlce